```
                    GenCore version 5.1.6
            Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:      July 21, 2004, 12:44:33 ; Search time 246 Seconds
                                        (without alignments)
                                        7697.124 Million cell updates/sec

Title:            US-09-270-437D-6
Perfect score:    3412
Sequence:         1 ggcagcggaggaggcgagga.........aacctgaaatgttattt 3412

Scoring table:    OLIGO_NUC
                  Gapop 60.0 , Gapext 60.0

Searched:         682709 seqs, 277475446 residues

Word size :       0

Total number of hits satisfying chosen parameters:      1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :     Issued Patents NA:*
        1: /cgn2_6/ptcdata/2/ina/5A_COMB.seq:*
        2: /cgn2_6/ptcdata/2/ina/5B_COMB.seq:*
        3: /cgn2_6/ptcdata/2/ina/6A_COMB.seq:*
        4: /cgn2_6/ptcdata/2/ina/6B_COMB.seq:*
        5: /cgn2_6/ptcdata/2/ina/PCTUS_COMB.seq:*
        6: /cgn2_6/ptcdata/2/ina/backfiles1.seq:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

                                SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3412	100.0	3412	3	US-09-061-709-6	Sequence 6, Appli
2	3412	100.0	3412	4	US-09-889-651-6	Sequence 6, Appli
3	2143	62.8	3283	3	US-09-061-709-8	Sequence 8, Appli
4	2143	62.8	3283	4	US-09-899-651-8	Sequence 8, Appli
c 5	34	1.0	317	4	US-09-621-976-13846	Sequence 13846, A
6	25	0.7	1149	4	US-09-644-460-36	Sequence 36, Appli
7	22	0.6	2224	4	US-09-261-885-1	Sequence 1, Appli
8	22	0.6	11049	3	US-10-204-708-23	Sequence 23, Appl
9	21	0.6	747	1	US-09-134-001C-2248	Sequence 2248, Ap
c 10	21	0.6	1602	1	US-08-075-533-1	Sequence 1, Appli
c 11	21	0.6	1602	2	US-08-948-176-1	Sequence 1, Appli
c 12	21	0.6	1602	5	PCT-US91-09160-1	Sequence 1, Appli
c 13	21	0.6	8961	4	US-10-204-708-80	Sequence 80, Appl
14	20	0.6	591	4	US-09-134-000C-2872	Sequence 2872, Ap
c 15	20	0.6	592	4	US-08-585-533A-27	Sequence 27, Appl
16	20	0.6	1539	4	US-09-364-230-29	Sequence 29, Appl
17	20	0.6	2199	4	US-09-708-725A-3	Sequence 3, Appli
c 18	20	0.6	2831	4	US-09-477-133A-117	Sequence 117, App
19	20	0.6	5428	4	US-09-625-972-21	Sequence 21, Appl
20	20	0.6	6070	4	US-10-204-708-10	Sequence 10, Appl
21	20	0.6	7174	4	US-08-961-527-189	Sequence 189, App
22	20	0.6	9641	4	US-09-625-972-22	Sequence 22, Appl
23	20	0.6	51259	3	US-08-781-891-209	Sequence 209, App
c 24	20	0.6	51259	4	US-09-618-166-209	Sequence 209, App
c 25	20	0.6	4403765	3	US-09-103-840A-1	Sequence 1, Appli
c 26	20	0.6	4411529	3	US-09-103-840A-1	Sequence 2, Appli
c 27	19	0.6	51	4	US-09-443-199C-203	Sequence 203, App

c 28	19	0.6	51	4	US-09-443-199C-204	Sequence 204, App
c 29	19	0.6	213	4	US-09-543-681A-1402	Sequence 1402, Ap
30	19	0.6	401	4	US-09-621-976-11383	Sequence 11383, A
31	19	0.6	1001	4	US-09-671-317-270	Sequence 270, App
32	19	0.6	1001	4	US-09-671-317-461	Sequence 461, App
33	19	0.6	1002	3	US-08-943-731-97	Sequence 97, Appl
34	19	0.6	1149	4	US-09-328-352-3312	Sequence 3312, Ap
35	19	0.6	1281	4	US-09-328-352-2290	Sequence 2290, Ap
c 36	19	0.6	1344	4	US-09-620-312D-937	Sequence 937, App
37	19	0.6	1443	2	US-09-802-927-2	Sequence 2, Appli
38	19	0.6	2186	1	US-08-878-546-9	Sequence 9, Appli
c 39	19	0.6	2238	1	US-08-173-436A-4	Sequence 4, Appli
c 40	19	0.6	2416	4	US-09-201-936-41	Sequence 41, Appl
c 41	19	0.6	2416	4	US-09-011-356-41	Sequence 41, Appl
c 42	19	0.6	2416	4	US-09-672-717-228	Sequence 228, App
c 43	19	0.6	2422	1	US-08-475-845-1	Sequence 1, Appli
c 44	19	0.6	2422	1	US-08-327-690-1	Sequence 1, Appli
c 45	19	0.6	2422	2	US-08-660-289-1	Sequence 1, Appli

```
                                ALIGNMENTS

RESULT 1
US-09-061-709-6
; Sequence 6, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 6
;   LENGTH: 3412
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-061-709-6

Query Match             100.0%;  Score 3412;  DB 3;  Length 3412;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 3412;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     1 GGCAGCGGAGGAGGCGAGGAGCGCGGGTACCGGGCGGGGAGCCGCGGCGCGGGCTCTCGGGG 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GGCAGCGGAGGAGGCGAGGAGCGCGGGTACCGGGCGGGGAGCCGCGGCGCGGGCTCTCGGGG 60

Qy    61 AAGAGACGGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCCGTCACCGCGAC 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 AAGAGACGGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCCGTCACCGCGAC 120

Qy   121 GACCTTCGGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGGGGGACCAGGTCCTGCTGAAG 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GACCTTCGGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGGGGGACCAGGTCCTGCTGAAG 180

Qy   181 TCCGGCTACGGCCTTCGTGGTGACTACCCCAGACTGGGCCATCCGGCCATCAGAGACC 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TCCGGCTACGGCCTTCGTGGTGACTACCCCAGACTGGGCCATCCGGCCATCAGAGACC 240

Qy   241 CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA 300

Qy   301 AAGTCAAGGAGCAGGAGAAATTCAGATTGAAACATCCCTCCTCACCTGCAGTGGGAGGTG 360
```

```
Db                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          301     AAGCTAAGGAGCAGGAAAATTCAGATTCGAAACATCCCTCCCTCACCTGCAGTGGGAGGTG  360

Qy                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          361     TTGGATGGACTTTTGGCTCAATATGGACAGTGGACAATGTGGAACAAGTCAACACAGAC    420
Db                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          361     TTTGGATGATGGACTTTTGGCTCAATATGGACAGTGGACAATGTGGAACAAGTCAACACAGAC  420

Qy                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          421     ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAGAGAGAGAAATAGCCATG        480
Db                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          421     ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAGAGAGAGAAATAGCCATG        480

Qy                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          481     GAGAAGCTAAGCGGGCATCAGTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGAT   540
Db                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          481     GAGAAGCTAAGCGGGCATCAGTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGAT   540

Qy                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          541     GAAGAGGTGAGCTCCCTTCGCCCCCTCAGCGAGCCCAGCCTGGGGACCACTCTTCCCGG    600
Db                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          541     GAAGAGGTGAGCTCCCTTCGCCCCCTCAGCGAGCCCAGCCTGGGGACCACTCTTCCCGG    600

Qy                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          601     GAGCAAGGCCACGCCCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGG   660
Db                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          601     GAGCAAGGCCACGCCCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGG   660

Qy                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          661     ATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAG    720
Db                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          661     ATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAG    720

Qy                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          721     AACATCACTAAGCAGACCAGTCCCGGGTAGATATCCATAGAAAAGAACTCTGGAGCT     780
Db                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          721     AACATCACTAAGCAGACCAGTCCCGGGTAGATATCCATAGAAAAGAACTCTGGAGCT     780

Qy                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          781     GCAGAGAAGCCTGTCACCATCCATGCACCCCAGAGGGACTTCTGAAGCTGCGCATG     840
Db                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          781     GCAGAGAAGCCTGTCACCATCCATGCACCCCAGAGGGACTTCTGAAGCTGCGCATG     840

Qy                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          841     ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG   900
Db                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          841     ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG   900

Qy                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          901     AAAATCTTGGCACACAATGGCTTGGTTGGAAGAAGAAGGCAGAAATTTG         960
Db                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          901     AAAATCTTGGCACACAATGGCTTGGTTGGAAGAAGAAGGCAGAAATTTG         960

Qy                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          961     AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020
Db                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          961     AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020

Qy                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         1021     ATATACAAACCCGGAAAGAACCATCACTGTGAGGGCACAGTTGAGGCCTGTGCCAGTGCT 1080
Db                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         1021     ATATACAAACCCGGAAAGAACCATCACTGTGAGGGCACAGTTGAGGCCTGTGCCAGTGCT 1080

Qy                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         1081     GAGATAGAGATTATGAAGAGCTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTAAC  1140
Db                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         1081     GAGATAGAGATTATGAAGAGCTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTAAC  1140

Qy                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         1141     CAACAAGCCAATCTGATCCCGGATTGAACCTGAACCTGGCATCTTTTCAACAGGA    1200
Db                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         1141     CCCTTCACTACCCACTCCGATATCTTCCACAGAGCAGGTTGAACCTGGCATCTTTTCAACAGGA  1200

Qy                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         1201     CTGTCCGTGCTATCTCCACCACTCACTCTTATCCCAGACGAGATTGAATCTTCATCCCAACCCACCAC 1260
Db                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         1201     CTGTCCGTGCTATCTCCACCACTCACTCTTATCCCAGACGAGATTGAATCTTCATCCCAACCCACCAC 1260

Qy                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         1261     CCCTTCACTACCCACTCCGATACTTCTTCCCAGCCTGTACCCCCGCTTGGCCCG      1320
Db                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         1261     CCCTTCACTACCCACTCCGATACTTCTTCCCAGCCTGTACCCCCGCTTGGCCCG      1320

Qy                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         1321     TTCCCGCATCATCATCTTCCACAGAGCAGAGAGATTGCAACTCTTATCCCAACCCAC    1380
Db                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         1321     TTCCCGCATCATCATCTTCCACAGAGCAGAGAGATTGCAACTCTTATCCCAACCCAC    1380

Qy                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         1381     GCTGTGGGCGCCATCATCGGAAGAAGAGGGCCACACATCAAACAGCTGCCGAGATTCGCC  1440
```

```
Db                GCTGTGGGCGCCATCATCGGAAGAAGAGGGCCACACATCAAACAGCTGCCGAGATTCGCC 1440

Qy                GGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGAGGCCGAAGGCGTCAGCGGGATGGTCATC 1500
Db                GGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGAGGCCGAAGGCGTCAGCGGGATGGTCATC 1500

Qy                ATCACCGGGCACCGGGACCGGCCGAAGCCAGTTCAAGGCTCAAGGGACGGATCTTTGGAAACTGAAA 1560
Db                ATCACCGGGCACCGGGACCGGCCGAAGCCAGTTCAAGGCTCAAGGGACGGATCTTTGGAAACTGAAA 1560

Qy                GAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGGAAGCGCATATCAGAGTGCCC 1620
Db                GAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGGAAGCGCATATCAGAGTGCCC 1620

Qy                TCTTCCACAGCTGCAGAAGTCATCGTGCCCTCGTGACCAACGCCAGATGAAAATGAGGAGAGTG 1680
Db                TCTTCCACAGCTGCAGAAGTCATCGTGCCCTGTGACCAACGCCAGATGAAAATGAGGAGAGTG 1680

Qy                TTAACCAGTGCAGAAGTCAGAATTATGGGCACTTCTTTGCTAGCCAGACTGCACAGCCCAAGATCAGGGAA 1740
Db                TTAACCAGTGCAGAAGTCATCGTGCCCTGTGACCAACGCCAGATGAAAATGAGGAGAGTG 1740

Qy                ATCGTCAGAATTATCGGCACTTCTTTGCTAGCCAGACTGCACAGCCCAAGATCAGGGAA 1800
Db                ATCGTCAGAATTATCGGCACTTCTTTGCTAGCCAGACTGCACAGCCCAAGATCAGGGAA 1800

Qy                ATTGTACAACAGGTGAAGCAGCAGCAGCAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGC 1860
Db                ATTGTACAACAGGTGAAGCAGCAGCAGCAGCAGAAATACCCTCAGGGAGTCGCGCTCACAGCGC 1860

Qy                AGCAAGTGAGGCTTCCCACAGGCAGCAGAACAACGATGAATGTAGCCCTTCCAACAC 1920
Db                AGCAAGTGAGGCTCGCGGAGGCGGCCGCAGCCAGCAGGATGAATGTAGCCCTTCCAACAC 1920

Qy                CTGACAGAATGAGACCAAAACGCCAGCGCAGCCGCAGGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGC 1980
Db                CTGACAGAATGAGACCAAACGCCAGCGCAGCCGCAGGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGC 1980

Qy                GAATGAGAAGTCTGCGGAGGCGGGCCGGAGGTCAGCGCCAGTTGCCAGAACCACCGCCAGTTTGCAGCC 2040
Db                GAATGAGAAGTCTGCGGAGGCGGGCCGGAGGTCAGCGCCAGTTGCCAGAACCACCGCCAGTTTGCAGCC 2040

Qy                CGAGGAGGGGCCGGGGAGGCGGGCCGGAGGTCAGCGCCAGGTCGCCGAGCCACCACCACCATCCATCCTCCTGAA 2100
Db                CGAGGAGGGGCCGGGGAGGCGGGCCGGAGGTCAGCGCCAGGTCGCCGAGCCACCACCACCATCCATCCTCCTGAA 2100

Qy                CCCCAGGGCTTCTGCAGCTTCAGCTTCAGCCTGTCCAGTGGTGTTCAAAGCCAAGC 2160
Db                CCCCAGGGCTTCTGCAGCTTCAGCTTCAGCCTGTCCAGTGGTGTTCAAAGCCAAGC 2160

Qy                CTCCCAGACGCTGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2220
Db                CTCCCAGACGCTGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2220

Qy                AAAATGCACACCCCTTTTCTCGTGTACATGTGTGTACAGGTGTGTACATATAGAAA 2280
Db                AAAATGCACACCCCTTTTCTCGTGTACATGTGTGTGTACAGGTGTGTACATATAGAAA 2280

Qy                GGGAAGATGTTAAGGATATGTGGCCTGTGGCTGTCCTCGCAGCGGTACAGGGTAATATAT 2340
Db                GGGAAGATGTTAAGGATATGTGGGTTACACAGGGTGCCTCGCAGCGGTAATATAT 2340

Qy                TTTTTCTTTTTAAAGAATAAGGCTTTTGGACTTTAAAGAATAAAGTCTTTGGGAG 2400
Db                TTTTTCTTTTTAAAGAATAAAGGCTTTTGGACTTTAAAGAATAAAGTCTTTGGGAG 2400

Qy                GTCTCACGGTGTGAGAGACGTGAGAGACGTTTGAGCCACACCGCACCCCACAAAATTCACCCGAGAGGGAAAT 2460
Db                GTCTCACGGTGTGAGAGACGTGAGAGACGTTTGAGCCACACCGCACCCCACAAAATTCACCCGAGAGGGAAAT 2520
```

; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
; TITLE OF INVENTION: Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/899,651
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/09/061,709
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8

SEQ ID NO 6
; LENGTH: 3412
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-899-651-6

Query Match   100.0%; Score 3412; DB 4; Length 3412;
Best Local Similarity   100.0%; Pred. No. 0;
Matches 3412; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy      1 GGCAGCGGAGGAGGCGGAGGAGCGCCGCGGTACCGGCCCGGGGGAGCCGCGGCTCTCGGGGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGCAGCGGAGGAGGCGGAGGAGCGCCGCGGTACCGGCCCGGGGGAGCCGCGGCTCTCGGGGG 60

Qy     61 AAGAGACGGATGATGAACAAGCTTACATCGGGAGACGTCGAGCCCCGCCGTCACCGCCGAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAGAGACGGATGATGAACAAGCTTACATCGGGAGACGTCGAGCCCCGCCGTCACCGCCGAC 120

Qy    121 GACCTTCCGGGACGAGCTTCTTTGGGGACAGGAGCTGCCCCTGCGGGACAGGTCCTGCTGAAG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GACCTTCCGGGACGAGCTTCTTTGGGGACAGGAGCTGCCCCTGCGGGACAGGTCCTGCTGAAG 180

Qy    181 TCCGGCTACCGCCTTCGTGGACTACCCGACCAGAACTGGGCCATCCGGCCATCGAGACC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCCGGCTACCGCCTTCGTGGACTACCCGACCAGAACTGGGCCATCCGGCCATCGAGACC 240

Qy    241 CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAACATCAGTTACTCAGTCTAAA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAACATCAGTTACTCAGTCTAAA 300

Qy    301 AAGCTAAGGAGCAGGAAAATTCAGATTCGGACGTGGAGAATGTGGACAAGTCAACACAGAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AAGCTAAGGAGCAGGAAAATTCAGATTCGGACGTGGAGAATGTGGACAAGTCAACACAGAC 360

Qy    361 TTGGATGGACTTTTGGCTCAATATGGACAGTGGACAGTTGTAAAGCTGATAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTGGATGGACTTTTGGCTCAATATGGACAGTGGACAGTTGTAAAGCTGATAG 420

Qy    421 ACAGAACCGCGTGTCAACGTCACATATGCAACAAGAGAGAGCAAAAATAGCCATG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ACAGAACCGCGTGTCAACGTCACATATGCAACAAGAGAGAGCAAAAATAGCCATG 480

Qy    481 GAGAGCTAAGCGGGCATCAGTTGAACTCAGTTTGAAGCGTGGGACCACTCTTCCCGG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GAGAGCTAAGCGGGCATCAGTTGAACTCAGTTTGAAGCGTGGGACCACTCTTCCCGG 540

Qy    541 GAAGAGGTGAGCTCCCCTTGCGCCCCTCGGGGACCCAGCGTGGGGACCACTCTTCCCGG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GAAGAGGTGAGCTCCCCTTGCGCCCCTCGGGGACCCAGCGTGGGGACCACTCTTCCCGG 600

Qy    601 GAGCAAGGCCACGCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GAGCAAGGCCACGCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGG 660

Qy    661 ATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAG 720

Qy    721 AACATCACTAAGCAGACCCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AACATCACTAAGCAGACCCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT 780
```

```
Qy   2521 CTCGTCGGAGGACACTCACGGCAGTTCGGATCACCTGTATGTCAACAGAAGGGATA 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 CTCGTCGGAGGACACTCACGGCAGTTCGGATCACCTGTATGTCAACAGAAGGGATA 2580

Qy   2581 CCGTCTCCTTGAAGAGGAAACTCTGCACTCCTGTCTCAGCTCATACACCCATT 2640
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 CCGTCTCCTTGAAGAGGAAACTCTGCACTCCTGTCTCAGCTCATACACCCATT 2640

Qy   2641 TCTCTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATAATTCTCTGTCT 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 TCTCTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATAATTCTCTGTCT 2700

Qy   2701 CTCTCGTGTTTATCTCTCCCCCTCCCCTTCTTCTCTCCCATCTCCATTCTTTGAA 2760
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 CTCTCGTGTTTATCTCTCCCCCTCCCCTTCTTCTCTCCCATCTCCATTCTTTGAA 2760

Qy   2761 TTTCTCATCCTCCATCTCCATCTCCCGTATCTACGGCACCCCCCCCCCAGGCAAAGCA 2820
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 TTTCTCATCCTCCATCTCCATCTCCCGTATCTACGGCACCCCCCCCCCAGGCAAAGCA 2820

Qy   2821 GTGCTCTGAGTATCACATCACACAAAGGAACAAAAGCGAAACACACAACCAGCCTCAA 2880
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 GTGCTCTGAGTATCACATCACACAAAGGAACAAAAGCGAAACACACAACCAGCCTCAA 2880

Qy   2881 CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAG 2940
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAG 2940

Qy   2941 AGGAAAACACAGAACCCACCAAACCAACCAATCAACCAAACAAAGAAAAATTCCACAATG 3000
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 AGGAAAACACAGAACCCACCAAACCAACCAATCAACCAAACAAAGAAAAATTCCACAATG 3000

Qy   3001 AAAGAATGTATTTTGTCTTTTGCATTTTGGTGTATAAGCCATCAATATTCAGCAAAATG 3060
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 AAAGAATGTATTTTGTCTTTTGCATTTTGGTGTATAAGCCATCAATATTCAGCAAAATG 3060

Qy   3061 ATTCCTTTCTTTAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCC 3120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 ATTCCTTTCTTTAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCC 3120

Qy   3121 CAGGGGCGTTAAATTCACAGATTTTTTTAACGAGAAAAACACACAGAGAGAGCTACCTCAG 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 CAGGGGCGTTAAATTCACAGATTTTTTTAACGAGAAAAACACACAGAGAGAGCTACCTCAG 3180

Qy   3181 GTGTTTTTACCTCAGCACCTTGCTCTCTGTGTTCCCTTAGAGATTTGTAAAGCTGATAG 3240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 GTGTTTTTACCTCAGCACCTTGCTCTCTGTGTTCCCTTAGAGATTTGTAAAGCTGATAG 3240

Qy   3241 TTGGAGCATTTTTTATTTTTTAATAAAATGAGTTGGAAAAAATAAGATATCAACT 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 TTGGAGCATTTTTTATTTTTTAATAAAATGAGTTGGAAAAAATAAGATATCAACT 3300

Qy   3301 GCCAGCCTGGAGAGGTGACAGTGCAACAGTCCAAGTGTGCAACAGCTGTGTCTTCTTCCGCT 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 GCCAGCCTGGAGAGGTGACAGTGCAACAGTCCAAGTGTGCAACAGCTGTGTCTTCTTCCGCT 3360

Qy   3361 AGCCAAGAACCNATATGGCCTTCTTTTGGACAAAACCTTGAAAATGTTTATTT 3412
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 AGCCAAGAACCNATATGGCCTTCTTTTGGACAAAACCTTGAAAATGTTTATTT 3412
```

RESULT 2
US-09-899-651-6
; Sequence 6, Application US/09899651
; Patent No. 6576756.
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander

```
Qy   781 GCAGAGAAGCCTGTCACCATCATGCACCCCAGAGGGGACTTCTGAAGCATGCCGCCATG 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GCAGAGAAGCCTGTCACCATCATGCACCCCAGAGGGGACTTCTGAAGCATGCCGCCATG 840

Qy   841 ATTCTTGAAATCATGCAGAAAGAGAGGCAGATGAGACCCGAAGAGATTCCTCTG 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 ATTCTTGAAATCATGCAGAAAGAGAGGCAGATGAGACCCGAAGAGATTCCTCTG 900

Qy   901 AAAATCTTGGCACAATGGGCTTGGTTGGAAGACTGATTGGAAAGAAGGCAGAAATTTG 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 AAAATCTTGGCACAATGGGCTTGGTTGGAAGACTGATTGGAAAGAAGGCAGAAATTTG 960

Qy   961 AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020

Qy  1021 ATATACAACCCGGAAAGAAGAACCATCACTGTGAAGGCACAGTTGAGGCCTGTGCCAGTGCT 1080
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 ATATACAACCCGGAAAGAAGAACCATCACTGTGAAGGCACAGTTGAGGCCTGTGCCAGTGCT 1080

Qy  1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGTGGCTGTTAAC 1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGTGGCTGTTAAC 1140

Qy  1141 CAACAAGCCAATCTGATCCCAGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGA 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 CAACAAGCCAATCTGATCCCAGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGA 1200

Qy  1201 CTGTCCGTGCTATCTCCACCAGCAGGCCCCGGAGCTCCCCGCTGCCCCCTACCAC 1260
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 CTGTCCGTGCTATCTCCACCAGCAGGCCCCGGAGCTCCCCGCTGCCCCCTACCAC 1260

Qy  1261 CCCTTCACTACCCCACTCCGGATACTTCTCCAGCAGGAGGAGATTGTGAATCTTCATCACCAGTTTGGCCCG 1320
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 CCCTTCACTACCCCACTCCGGATACTTCTCCAGCAGGAGGAGATTGTGAATCTTCATCACCAGTTTGGCCCG 1320

Qy  1321 TTCCCGCATCATCACTCTTATCCAGACGAGGAGATTGTGAATCTTCATCCCAACCCAG 1380
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 TTCCCGCATCATCACTCTTATCCAGACGAGGAGATTGTGAATCTTCATCCCAACCCAG 1380

Qy  1381 GCTGTGTGGGCCCATCATGGGGAAGAAGGGGGCCACACATCAAACAGTCTGGCGAGATTCGCC 1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 GCTGTGTGGGCCCATCATGGGGAAGAAGGGGGCCACACATCAAACAGTCTGGCGAGATTCGCC 1440

Qy  1441 GGAGCCTCTATCAAGATTGCCCTGCGGAAGCCCAGAGCTCAGCGAAAGGATGGTCATC 1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 GGAGCCTCTATCAAGATTGCCCTGCGGAAGCCCAGAGCTCAGCGAAAGGATGGTCATC 1500

Qy  1501 ATCACGGGCCACCGGAAGCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAA 1560
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 ATCACGGGCCACCGGAAGCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAA 1560

Qy  1561 GAGGAAAACTTCTTTAACCCAAAGAAGAAGTGAAGCTGGAAGCGCATATCAGAGTGCCC 1620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 GAGGAAAACTTCTTTAACCCAAAGAAGAAGTGAAGCTGGAAGCGCATATCAGAGTGCCC 1620

Qy  1621 TCTTCCACAGCTGGCCGGGTGATTGGCAAGGTGCAAGACCGTGAACTGCAGAAC 1680
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621 TCTTCCACAGCTGGCCGGGTGATTGGCAAGGTGCAAGACCGTGAACTGCAGAAC 1680

Qy  1681 TTAACCAGTTGCAGAAGTCATCGTGACCCAGTCGTGACCAACAAACGCCAGATGAAAATGAGGAAGTG 1740
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681 TTAACCAGTTGCAGAAGTCATCGTGACCCAGTCGTGACCAACAAACGCCAGATGAAAATGAGGAAGTG 1740

Qy  1741 ATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCGAGTCGCACGCCTCAGCAGCGC 1800
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741 ATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCGAGTCGCACGCCTCAGCAGCGC 1800

Qy  1801 ATTGTACAACAGGTTGACTCAAGCAGTGAAGCGCCTCAGGGAGTCGCCTCACAGCGC 1860
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801 ATTGTACAACAGGTTGACTCAAGCAGTGAAGCGCCTCAGGGAGTCGCCTCACAGCGC 1860

Qy  1861 AGCAAGTGAGGCTCCCACAGGCACCACGAGCTCCTTCCAACAC 1920
         |||||||||||||||||||||||||||||||||||||||||||
Db  1861 AGCAAGTGAGGCTCCCACAGGCACCACGAGCTCCTTCCAACAC 1920

Qy  1921 CTGACAGAATGAGACCAGCCAGCCAGATGAGGAGCAAACCAAAGACCATCTGAG 1980
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921 CTGACAGAATGAGACCAGCCAGCCAGATGAGGAGCAAACCAAAGACCATCTGAG 1980

Qy  1981 GAATGAGAAGTCTGCGGAGCGGCGCCAGGGACTCTGCCGAGCGCCCTGAGCAGACCCCGAGGGC 2040
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981 GAATGAGAAGTCTGCGGAGCGGCGCCAGGGACTCTGCCGAGCGCCCTGAGACCCCCAGGGGC 2040

Qy  2041 CGAGGAGGGGCGGCGGGAAGTCAGCCAGTTTGCCAGAACCACCGAGCCCGCTCCCGCCC 2100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041 CGAGGAGGGGCGGCGGGAAGTCAGCCAGTTTGCCAGAACCACCGAGCCCGCTCCCGCC 2100

Qy  2101 CCCCAGGGCTTCTGCAGCCTTCAGCCATCCACTTCACCATCCACTTCACCATCCACTCCTCCTGAA 2160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101 CCCCAGGGCTTCTGCAGCCTTCAGCCATCCACTTCACCATCCACTTCACCATCCACTTCTCCTGAA 2160

Qy  2161 CTCCCACGACGCTATCCCTTTTCTGTGCAAATCGTCTCGTGTACATGTGTGTGTTCAAAGCCAAGC 2220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161 CTCCCACGACGCTATCCCTTTTCTGTGCAAATCGTCTCGTGTACATGTGTGTGTTCAAAGCCAAGC 2220

Qy  2221 AAAATGCACACCCTTTTCTGTGCAAATCGTCTCGTGTACATGTGTGTACATTATTAGAAA 2280
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221 AAAATGCACACCCTTTTCTGTGCAAATCGTCTCGTGTACATGTGTGTACATTATTAGAAA 2280

Qy  2281 GGGAAGATGTTAAGATATGTGGCCTGTGTGGGTTACACAGGGTGCCTGCAGCCGGTAATATAT 2340
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281 GGGAAGATGTTAAGATATGTGGCCTGTGTGGGTTACACAGGGTGCCTGCAGCCGGTAATATAT 2340

Qy  2341 TTTAGAAATAATATATCAAATAACTCAATTTTAATCAATTATTAATTTT 2400
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341 TTTAGAAATAATATATCAAATAACTCAATTTTAATCAATTATTAATTTT 2400

Qy  2401 TTTTCTTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG 2460
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401 TTTTCTTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG 2460

Qy  2461 GTCTCACGGTGTAGAGAGCAGCTTTGAGCCACCCGCACAAAAATTCACCCAGAGGGAAAT 2520
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461 GTCTCACGGTGTAGAGAGCAGCTTTGAGCCACCCGCACAAAAATTCACCCAGAGGGAAAT 2520

Qy  2521 CTCGTCGGAAGGACACTCACGGCAGTTCTGATCACGTGTATGTCAACGACGAAGGGATA 2580
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521 CTCGTCGGAAGGACACTCACGGCAGTTCTGATCACGTGTATGTCAACGACGAAGGGATA 2580

Qy  2581 CCGTCTCCTTGAAGAGGAAACTCTGTCCTCAGTTGCCTGTCTCAGCTGCTCAGCTCTCATCACACCCATT 2640
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581 CCGTCTCCTTGAAGAGGAAACTCTGTCCTCAGTTGCCTGTCTCAGCTGCTCAGCTCTCATCACACCCATT 2640

Qy  2641 TCTCTTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATATTGCTATATAATTCTCTGTCT 2700
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641 TCTCTTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATATTGCTATATAATTCTCTGTCT 2700

Qy  2701 CTCTTCGTTTATCTCTCCCATCTCAATCCCGTCATCCTTCTTCCTCATTCTTTGAA 2760
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701 CTCTTCGTTTATCTCTCCCATCTCAATCCCGTCATCCTTCTTCCTCATTCTTTGAA 2760

Qy  2761 TTTCCTCATCCCTCAATCCCGTATCTCACCGCACCCCCCCCGTAATCTTACGGACAAAGCA 2820
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761 TTTCCTCATCCCTCAATCCCGTATCTCACCGCACCCCCCCCGTATCTTACGGACAAAGCA 2820

Qy  2821 GTGCTCTGAGTATCACATCACAAAAGGAACACAAAGCGAAAACCACACAAACCAGCCTCAA 2880
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821 GTGCTCTGAGTATCACATCACAAAAGGAACACAAAGCGAAAACCACACAAACCAGCCTCAA 2880

Qy  2881 CTTACACTTGGTTACTCCAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGTTGGAAG 2940
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881 CTTACACTTGGTTACTCCAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGTTGGAAG 2940

Qy  2941 AGGAAAACAGGAACCCACCAAACCAACCAATCAACCAACCAAAGAAAATTCCACAATG 3000
```

Db 2941 AGGAAAACAGGAACCCACCAACCAATCAACCAAACAAGAGAAAAATTCCACAATG 3000
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 ATCAAGATTGCCCCTGCGAAGGCCCAGACGTCAGCGAAGGATGGTCATCATCACCGGG 1380

Qy 3001 AAAGAATGTATTTGTCTTTTGCATTTGTGTATAAGCCATCAATATTCAGCAAAAATG 3060
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1510 CCACCGGAAGCCAGTTCAAGGCCCAGGCCCAGGACGGATCTTTGGGAAACTGAAAGAGAGGAAAAC 1569

Db 3001 AAAGAATGTATTTGTCTTTTGCATTTGTGTATAAGCCATCAATATTCAGCAAAAATG 3060
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1381 CCACCGGAAGCCAGTTCAAGGCCCAGGCCCAGGACGGATCTTTGGGAAACTGAAAGAGAGGAAAAC 1440

Qy 3061 ATTCCTTTCTTAAAAAAAAATGTGGAGAAAGTAGAAAATTTACCAAGGTTGTTGGCC 3120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1570 TTCTTTAACCCAAAGAAGAAGTGGAAGCTGGAAGCCGCATATCAGAGTGCCCTCTTCCACA 1629

Db 3061 ATTCCTTTCTTAAAAAAAAATGTGGAGAAAGTAGAAAATTTACCAAGGTTGTTGGCC 3120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1441 TTCTTTAACCCAAAGAAGAAGTGGAAGCTGGAAGCCGCATATCAGAGTGCCCTCTTCCACA 1500

Qy 3121 CAGGGGGTTAAATTCACAGATTTTTTAACAGAAAAACACAGAAGAGAGCTACCTCAG 3180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1630 GCTGCGCCGGTGATTGGCAAAGGTGGCAAGACCGTGAACGACGAACTGCAGAACTAACCAGT 1689

Db 3121 CAGGGGGTTAAATTCACAGATTTTTTAACAGAAAAACACAGAAGAGAGCTACCTCAG 3180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1501 GCTGCGCCGGTGATTGGCAAAGGTGGCAAGACCGTGAACGACGAACTGCAGAACTAACCAGT 1560

Qy 3181 GTGTTTTTACCTCAGCACCTTGCTCTGTGTTTCCCTTAGAGATTTGTAAAGCTGATAG 3240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1690 GCAGAAGTCATCGTGCCTCGTGACCAGATGAAATGAGGAAATGAGGAAGTGATCGTCAGA 1749

Db 3181 GTGTTTTTACCTCAGCACCTTGCTCTGTGTTTCCCTTAGAGATTTGTAAAGCTGATAG 3240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1561 GCAGAAGTCATCGTGCCTCGTGACCAGATGAAATGAGGAAATGAGGAAGTGATCGTCAGA 1620

Qy 3241 TTGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACT 3300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1750 ATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGAGAAATTGTACAA 1809

Db 3241 TTGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACT 3300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1621 ATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGAGAAATTGTACAA 1680

Qy 3301 GCCAGCCTGGAGAAGTGACAGTCCAAGTGTGCAACAGCAGTGTTCTGAATTGTCTTCCGCT 3360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1810 CAGGTGAAGCAGGAGCAGGAGAAATACCCTCAGGGAGTGCGCCTCACAGCGCAGCAAGTGA 1869

Db 3301 GCCAGCCTGGAGAAGTGACAGTCCAAGTGTGCAACAGCAGTGTTCTGAATTGTCTTCCGCT 3360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1681 CAGGTGAAGCAGGAGCAGGAGAAATACCCTCAGGGAGTGCGCCTCACAGCGCAGCAAGTGA 1740

Qy 3361 AGCCAAGAACCNATATGGCCTTCTTTTGGACAAAATGTTTATTT 3412
         ||||||||||||||||||||||||||||||||||||||||||||
Qy 1870 GGCTCCCACAGGCACCAGCAAACAACGGATGAATGTAGCCCTTCCAACACCTGACAGAA 1929

Db 3361 AGCCAAGAACCNATATGGCCTTCTTTTGGACAAAATGTTTATTT 3412
         ||||||||||||||||||||||||||||||||||||||||||||
Db 1741 GGCTCCCACAGGCACCAGCAAACAACGGATGAATGTAGCCCTTCCAACACCTGACAGAA 1800

Qy 1930 TGAGACCAAACGCAGCAGCGAGGCGGCAGCAAACCATCGAAGACCATCTGAGGAATGAGAA 1989
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1801 TGAGACCAAACGCAGCAGCGAGGCGGCAGCAAACCAAAGACCATCTGAGGAATGAGAA 1860

Qy 1990 GTCTGCGGAGGCGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCGAGGGGGCCGAGGAGGG 2049
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1861 GTCTGCGGAGGCGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCGAGGGGGCCGAGGAGGG 1920

Qy 2050 GCGGGGAAGTCAGCCAGGTTGCCAGAACCACCGACCACCGAGCCCCGCCTCCCGGGGGC 2109
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1921 GCGGGGAAGTCAGCCAGGTTGCCAGAACCACCGACCACCGAGCCCCGCCTCCCGGGGGC 1980

Qy 2110 TTCTGCAGGCTTCAGCCATCCACCTTCACCATCGAGATCTCTCCACTCCCACGA 2169
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1981 TTCTGCAGGCTTCAGCCATCCACCTTCACCATCGAGATCTCTCCACTCCCACGA 2040

Qy 2170 CGCATCCCTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGCAAAATGCAC 2229
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2041 CGCTATCCCTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGCAAAATGCAC 2100

Qy 2230 ACCCTTTTCTGTGGCAAATCGTCTCGTACATGTGTGTATACATTAGAAAGGAAGATG 2289
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2101 ACCCTTTTCTGTGGCAAATCGTCTCGTACATGTGTGTATACATTAGAAAGGGAAGATG 2160

Qy 2290 TTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGCGGTAATATTTTAGAAAT 2349
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2161 TTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGCGGTAATATATTTTAGAAAT 2220

Qy 2350 AATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTATTTTTTTTCTTT 2409
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2221 AATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTATTTTTTTTCTTT 2280

Qy 2410 TTAAAGAGAAAGCAGGCTTGAGGCCACCCGACAAAATTCACCCAGAGGGAAATCTCGTCGGA 2469
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2281 TTAAAGAGAAAGCAGGCTTTGAGGCCACCCGACAAAATTCACCCAGAGGGAAATCTCGTCGGA 2340

Qy 2470 TGTGAGGAGGAGCTTTGAGGCGATTCGCCGGAGATTCGCCGGAGCCTCT 2529
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db 2341 TGTGAGGAGGAGCTTTGAGGCGACCCGACAAAATTCACCGAGATTCGCCGGAGCCTCT 2400

Qy 2530 AGGACACTCACGGACAGTTCTGGGATCACCTGTGTATGTCAACAGAGAAGGGGATTCGTCTCCT 2589
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 3
US-09-061-709-8
; Sequence 8, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
;   APPLICANT: Chen, Yao-Tseng
;   APPLICANT: Gure, Ali
;   APPLICANT: Tsang, Solam
;   APPLICANT: Stockert, Elisabeth
;   APPLICANT: Jager, Elke
;   APPLICANT: Knuth, Alexander
;   APPLICANT: Old, Lloyd J.
;   TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
;   TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
;   FILE REFERENCE: LUD 5538
;   CURRENT APPLICATION NUMBER: US/09/061,709B
;   CURRENT FILING DATE: 1998-04-17
;   NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 8
;   LENGTH: 3283
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-061-709-8

Query Match   62.8%; Score 2143; DB 3; Length 3283;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1270 ACCCACTCCGGATACTTCTCCAGACGAGATTGGCCCGTTGGCCCGTTCCCGCAT 1329
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 ACCCACTCCGGATACTTCTCCAGACGAGATTGGCCCGTTGGCCCGTTCCCGCAT 1200

Qy 1330 CATCACTCTTATCCAGAGCAGGAGAGATTGTGAATCTCTTCAGACAACGCTGGGC 1389
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 CATCACTCTTATCCAGAGCAGGAGAGATTGTGAATCTCTTCAGACAACGCTGTGGGC 1260

Qy 1390 GCCATCATCGGGAAGAAGGGGGCACACATCAAACAGCTGGCCGGAGATTCGCCGGAGCCTCT 1449
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 GCCATCATCGGGAAGAAGGGGGCCACACATCAAACAGCTGGCCGGAGATTCGCCGGAGCCTCT 1320

Qy 1450 ATCAAGATTGCCCCTGCGAAGGCCCAGACGTCAGCGAAGGATGGTCATCATCACCGGG 1509
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
;  TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
;  FILE REFERENCE: LUD 5538
;  CURRENT APPLICATION NUMBER: US/09/899,651
;  CURRENT FILING DATE: 2001-07-06
;  PRIOR APPLICATION NUMBER: US/09/061,709
;  PRIOR FILING DATE: 1998-04-17
;  NUMBER OF SEQ ID NOS: 8
SEQ ID NO 8
;  LENGTH: 3283
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
US-09-899-651-8

Query Match        62.8%;   Score 2143;   DB 4;   Length 3283;
Best Local Similarity  100.0%;  Pred. No. 0;
Matches 2143;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   1270 ACCCACTCCGATACTTCTCCAGCCTGTACCCCCATCACCAGTTGGCCCGTTCCCGCAT 1329
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 ACCCACTCCGATACTTCTCCAGCCTGTACCCCCATCACCAGTTGGCCCGTTCCCGCAT 1200

Qy   1330 CATCACTCCTATCCAGAGCAGGAGATGTGAATCTCTTCATCCCAACCAGGCTGTGGGC 1389
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CATCACTCCTATCCAGAGCAGGAGATGTGAATCTCTTCATCCCAACCAGGCTGTGGGC 1260

Qy   1390 GCCATCATCGGGAAGAAGAGGGGGCACACATCAAACAGCTGCGGAGATTCGCCGGAGCTCT 1449
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 GCCATCATCGGGAAGAAGAGGGGGCACACATCAAACAGCTGCGGAGATTCGCCGGAGCCTCT 1320

Qy   1450 ATCAAGATTGCCCCTGCGGAAGCCCAGAACGTCAGCGAGAGCGTCATCATCACCGGG 1509
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 ATCAAGATTGCCCCTGCGGAAGCCCAGAACGTCAGCGAGAGCGTCATCATCACCGGG 1380

Qy   1510 CCACCGGAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAAGAGGAAAAC 1569
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 CCACCGGAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAAGAGGAAAAC 1440

Qy   1570 TTCTTTAACCCAAAGAAGAAGAGTGAGCTGAAGCGCATATCAGAGTGCCCTTCTTCCACA 1629
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 TTCTTTAACCCAAAGAAGAAGAGTGAGCTGAAGCGCATATCAGAGTGCCCTTCTTCCACA 1500

Qy   1630 GCTGGCCGGGTGATTGGCAAGAGGTGGCAAGACCGTGAACGAACTGCAGAACTTAACCAGT 1689
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GCTGGCCGGGTGATTGGCAAGAGGTGGCAAGACCGTGAACGAACTGCAGAACTTAACCAGT 1560

Qy   1690 GCAGAAGTCATCGTGCCTGCCTCGTTGACCAAACGCCAGATGAAAATGAGGAAGTCGTCAGA 1749
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 GCAGAAGTCATCGTGCCTGCCTCGTTGACCAAACGCCAGATGAAAATGAGGAAGTCGTCAGA 1620

Qy   1750 ATTATCGGGCACTTCTTTGCTAGCCAGCAGAAATACCCTCAGGAGGTCAGCGGAGAATTGTACAA 1809
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ATTATCGGGCACTTCTTTGCTAGCCAGCAGAAATACCCTCAGGAGGTCAGCGGAGAATTGTACAA 1680

Qy   1810 CAGGTGAAGCAGGAGCAGAAATACCCTTCCAACACCTGACAGAA 1869
          |||||||||||||||||||||||||||||||||||||||||||
Db   1681 CAGGTGAAGCAGGAGCAGAAATACCCTTCCAACACCTGACAGAA 1740

Qy   1870 GGCTCCCAGGCACCACGGCACCAGGATGAATGTAGCCCTTCCAACACCTGACAGAA 1929
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GGCTCCCAGGCACCACGGCACCAGGATGAATGTAGCCCTTCCAACACCTGACAGAA 1800

Qy   1930 TGAGACCAAACGCAGCCAGCCAGATCGGAGCAAAGACCATCTGAGGAATGAGAA 1989
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TGAGACCAAACGCAGCCAGCCAGATCGGAGCAAAGACCATCTGAGGAATGAGAA 1860

Qy   1990 GTCTGCGGAGGCGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCGAGGAGGG 2049
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GTCTGCGGCGGCGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCGAGGAGGG 1920

Qy   2050 GCGGGGAAGGTCAGCCAGGTTTGCCAGAACCACCGGCCCCCGCCTCCGAGCCCCAGGGGC 2109
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 GCGGGGAAGGTCAGCCAGGTTTGCCAGAACCACCGGCCCCGCCTCCGAGCCCCAGGGGC 1980

Db   2401 AGGAACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAAGGGATACCGTCTCCT 2460

Qy   2590 TGAAGAGGAAACTCTGTCACTCCTCAGCCTGTTCATCACACCCATTTCTCTTTGC 2649
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 TGAAGAGGAAACTCTGTCACTCCTCAGCCTGTTCATCACACCCATTTCTCTTTGC 2520

Qy   2650 TTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCGTCTCTCCTCGTT 2709
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 TTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCGTCTCTCCTCGTT 2580

Qy   2710 TATCTCTCCCCCCCCCCCCCCCCTTCTTCCCATCTCCATCTCCATTCTTTGAATTTCCTCAT 2769
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 TATCTCTCCCCCTCCCCCCTTCTTCCCATCTCCATCTCCATTCTTTGAATTTCCTCAT 2640

Qy   2770 CCCTCCATCTCAATCCCGTATCTACGGCACCCCCCCCCCCAGGCAAAGCAGTGCTCTGA 2829
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 CCCTCCATCTCAATCCCGTATCTACGGCACCCCCCCCCCCAGGCAAAGCAGTGCTCTGA 2700

Qy   2830 GTATCACATCACACAAAGGAACAAAAGCGAAACACACAGCCTCAAACTTACACTT 2889
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 GTATCACATCACACAAAGGAACAAAAGCGAAACACACAGCCTCAAACTTACACTT 2760

Qy   2890 GGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAGAGGAAAACA 2949
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 GGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAGAGGAAAACA 2820

Qy   2950 GGAACCCACCAAACCAATCAACCAAACAAAGAAAAATTCCACAATGAAAGAATGT 3009
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 GGAACCCACCAAACCAATCAACCAAACAAAGAAAAATTCCACAATGAAAGAATGT 2880

Qy   3010 ATTTTGTCTTTTTGCATTTTGGTGTATAAGCCATCAATATTATTCAGCAAAATGATTCCTTTC 3069
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 ATTTTGTCTTTTTGCATTTTGGTGTATAAGCCATCAATATTATTCAGCAAAATGATTCCTTTC 2940

Qy   3070 TTTAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCCCAGGGCGTT 3129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 TTTAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCCCAGGGCGTT 3000

Qy   3130 AAATTCACAGATTTTTTTAACGAGAAAAACACAGAGAAGAAGCTACCTCAGGTGTTTTA 3189
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 AAATTCACAGATTTTTTTAACGAGAAAAACACAGAGAAGAAGCTACCTCAGGTGTTTTA 3060

Qy   3190 CCTCAGCACCTTGCTCTGTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGGAGCAT 3249
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 CCTCAGCACCTTGCTCTGTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGGAGCAT 3120

Qy   3250 TTTTTATTTTTTAATAAAAATGAGTTGAAAAAATTAAGATATCAACTGCCAGCCTG 3309
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 TTTTTATTTTTTAATAAAAATGAGTTGAAAAAATTAAGATATCAACTGCCAGCCTG 3180

Qy   3310 GAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCCGTAGCCAAGAA 3369
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 GAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCCGTAGCCAAGAA 3240

Qy   3370 CCNATATGGCCTTCTTTTTGGACAAACCTTGAAAATGTTTATTT 3412
          |||||||||||||||||||||||||||||||||||||||||||
Db   3241 CCNATATGGCCTTCTTTTTGGACAAACCTTGAAAATGTTTATTT 3283
```

RESULT 4
US-09-899-651-8
; Sequence 8, Application US/09899651
; Patent No. 6576756
; GENERAL INFORMATION:
APPLICANT: Chen, Yao-Tseng
APPLICANT: Gure, Ali
APPLICANT: Tsang, Solam
APPLICANT: Stockert, Elisabeth
APPLICANT: Jager, Elke
APPLICANT: Knuth, Alexander
APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
; TITLE OF INVENTION: Associated

```
Qy    3190  CCTCAGCACCTTGCTCTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGGAGCAT  3249
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3061  CCTCAGCACCTTGCTCTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGGAGCAT  3120

Qy    3250  TTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCAGCCTG  3309
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3121  TTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCAGCCTG  3180

Qy    3310  GAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTTCCGCTAGCCAAGAA  3369
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3181  GAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTTCCGAATTGTCTTCCGCTAGCCAAGAA  3240

Qy    3370  CCNATATGCCTTCTTTGGACAAACCTTGAAAATGTTTATTT  3412
            ||||||||||||||||||||||||||||||||||||
Db    3241  CCNATATGCCTTCTTTTTGGACAAACCTTGAAAATGTTTATTT  3283
```

RESULT 5
US-09-621-976-13846/c
; Sequence 13846, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 13846
; LENGTH: 317
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-13846

  Query Match   1.0%; Score 34; DB 4; Length 317;
  Best Local Similarity 100.0%; Pred. No. 3.3e-06;
  Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    2595  AGGAAACTCTGTCACTCCTCATGCCTGTCTAGCT  2628
            ||||||||||||||||||||||||||||||||||
Db     216  AGGAAACTCTGTCACTCCTCATGCCTGTCTAGCT  183
```

RESULT 6
US-09-644-460-36
; Sequence 36, Application US/09644460
; Patent No. 6657053
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; TITLE OF INVENTION: Reciprocal Subtraction Differential
; TITLE OF INVENTION: Display
; FILE REFERENCE: 34587-C-PCT-USA
; CURRENT APPLICATION NUMBER: US/09/644,460
; CURRENT FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: PCT/US99/04323
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: US 09/197,889
; PRIOR FILING DATE: 1998-11-23
; PRIOR APPLICATION NUMBER: US 09/185,115
; PRIOR FILING DATE: 1998-11-03
; PRIOR APPLICATION NUMBER: US 09/032,684
; PRIOR FILING DATE: 1998-02-27
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 1149
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-644-460-36

```
Qy    2110  TTCTGCAGGCTTCAGCCATCCACTTCACCATCCACTCGGATCTCCTCGAACTCCCACGA  2169
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981  TTCTGCAGGCTTCAGCCATCCACCATCCACTTCACCATCCACTCGGATCTCCTCGAACTCCCACGA  2040

Qy    2170  CGCTATCCCTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGCAAAATGCAC  2229
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041  CGCTATCCCTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGCAAAATGCAC  2100

Qy    2230  ACCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAAGGGAAGATG  2289
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101  ACCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAAGGGAAGATG  2160

Qy    2290  TTAAGATATGTGGCCTGTGGGTTACACAGGTGCCGCGGGTAATATATTTAGAAAT  2349
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2161  TTAAGATATGTGGCCTGTGGGTTACACAGGTGCCGCGGGTAATATATTTTTAGAAAT  2220

Qy    2350  AATATATCAAATAACTAACTCCAATTTTTTTAATCAATTATTAATTAATTTTTTTTTCTTT  2409
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2221  AATATATCAAATAACTAACTCAACTAACTCCAATTTTTTAATCAATTATTAATTAATTTTTTTTTCTTT  2280

Qy    2410  TTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTCACCG  2469
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2281  TTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTCACCG  2340

Qy    2470  TGTAGAGAGGAGCTTGAGGCACCCGCACAAAATTCACCCAGAGGGAAATCTCGTCGGA  2529
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2341  TGTAGAGAGGAGCTTGAGGCACCCGCACCACCCGCACAAAATTCACCCAGAGGGAAATCTCGTCGGA  2400

Qy    2530  AGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAGAGGATACCGTCTCCT  2589
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2401  AGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAGAGGATACCGTCTCCT  2460

Qy    2590  TGAAGAGGAAACTCTGTCACTCCCATGCCTGTCTAGCCCATACACCCATTTCTCTTTGC  2649
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2461  TGAAGAGGAAACTCTGTCACTCCCATGCCTGTCTAGCCCATACACCCATTTCTCTTTGC  2520

Qy    2650  TTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATATAATTCCTGTCTCCTCTCTGTT  2709
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2521  TTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATATAATTCCTGTCTCCTCTCTGTT  2580

Qy    2710  TATCTCTCCCCTCCCCCTCCCCCCTTCTTCTCCATCTCCATTCTTTGAATTCCTCAT  2769
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2581  TATCTCTCCCCTCCCCCTCCCCCCTTCTTCTCCATCTCCATTCTTTGAATTCCTCAT  2640

Qy    2770  CCCTCCATCTCAATCCCGTATCTACGGCACCCCCCCCCCCCCCAGGCAAAGCAGTGCTCTGA  2829
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2641  CCCTCCATCTCAATCCCGTATCTACGGCACCCACCCCCCCCCCCCCAGGCAAAGCAGTGCTCTGA  2700

Qy    2830  GTATCACATCACACAAAAGGAACAAAAGGGAAAACACACAACCAGCCTCAACTTACACTT  2889
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2701  GTATCACATCACACAAAAGGAACAAAAGGAACAAAAGGGAAAACACAACCAGCCTCAACTTACACTT  2760

Qy    2890  GGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGAGAAAACA  2949
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2761  GGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGAGAAAACA  2820

Qy    2950  GGAACCCACCAAACCAACCAACCAATCAACCAACAAAGAAAATTCCACAATGAAAGAATGT  3009
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2821  GGAACCCACCAAACCAACCAACCAATCAACCAACAAAGAAAATTCCACAATGAAAGAATGT  2880

Qy    3010  ATTTTGTCTTTTTGCATTTGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTC  3069
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2881  ATTTTGTCTTTTTGCATTTGTGTATAAGCCATCATCAATATTCAGCAAAATGATTCCTTTC  2940

Qy    3070  TTTAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGGCCCAGGGCGTT  3129
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2941  TTTAAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGGCCCAGGGCGTT  3000

Qy    3130  AAATTCACAGATTTTTTAACGAGAAAAAACACACAAGAAGAGCTACCTCCAGGTGTTTTTA  3189
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3001  AAATTCACAGATTTTTTTAACGAGAAAAAACACACAAGAAGAGCTACCTCCAGGTGTTTTTA  3060
```

```
Query Match             0.7%;  Score 25;  DB 4;  Length 1149;
Best Local Similarity  100.0%;  Pred. No. 0.066;
Matches   25;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy       2957 ACCAAACCACCAACCAATCAACCAAACA 2981
              |||||||||||||||||||||||||||||
Db        932 ACCAAACCACCAACCAATCAACCAAACA 956

RESULT 7
US-09-261-855-1
; Sequence 1, Application US/09261855A
; Patent No. 6255055
; GENERAL INFORMATION:
;   APPLICANT: Ross, Jeffrey
;   TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN
;   TITLE OF INVENTION: (CRD-BP) AND ITS NUCLEIC ACID SEQUENCE
;   FILE REFERENCE: 960296.95131
;   CURRENT APPLICATION NUMBER: US/09/261,855A
;   CURRENT FILING DATE: 1999-03-03
;   NUMBER OF SEQ ID NOS: 46
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2224
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-261-855-1

Query Match             0.6%;  Score 22;  DB 3;  Length 2224;
Best Local Similarity  100.0%;  Pred. No. 1.8;
Matches   22;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy       1376 CCCAGGCTGTGGGCGCCATCAT 1397
              ||||||||||||||||||||||
Db       1371 CCCAGGCTGTGGGCGCCATCAT 1392

RESULT 8
US-10-204-708-23
; Sequence 23, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;   APPLICANT: OLEK, Alexander
;   APPLICANT: PIEPENBROCK, Christian
;   APPLICANT: BERLIN, Kurt
;   TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;   TITLE OF INVENTION: by Assessing DNA Methylation
;   FILE REFERENCE: 5013 1012
;   CURRENT APPLICATION NUMBER: US/10/204,708
;   CURRENT FILING DATE: 2003-05-06
;   PRIOR APPLICATION NUMBER: PCT/EP01/03971
;   PRIOR FILING DATE: 2001-04-06
;   PRIOR APPLICATION NUMBER: DE 10019058.8
;   PRIOR FILING DATE: 2000-04-06
;   PRIOR APPLICATION NUMBER: DE 10019173.8
;   PRIOR FILING DATE: 2000-04-07
;   PRIOR APPLICATION NUMBER: DE 10032529.7
;   PRIOR FILING DATE: 2000-06-30
;   PRIOR APPLICATION NUMBER: DE 10043826.1
;   PRIOR FILING DATE: 2000-09-01
;   NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 23
;   LENGTH: 11049
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-23

Query Match             0.6%;  Score 22;  DB 4;  Length 11049;
Best Local Similarity  100.0%;  Pred. No. 1.8;
Matches   22;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Qy       3248 ATTTTTTATTTTTTAATAAA 3269
              |||||||||||||||||||
Db        680 ATTTTTTATTTTTTAATAAA 701

RESULT 9
US-09-134-001C-2248/c
; Sequence 2248, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;   APPLICANT: Lynn Doucette-Stamm et al
;   TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;   TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;   FILE REFERENCE: GTC-007
;   CURRENT APPLICATION NUMBER: US/09/134,001C
;   CURRENT FILING DATE: 1998-08-13
;   PRIOR APPLICATION NUMBER: US 60/064,964
;   PRIOR FILING DATE: 1997-11-08
;   PRIOR APPLICATION NUMBER: US 60/055,779
;   PRIOR FILING DATE: 1997-08-14
;   NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2248
;   LENGTH: 747
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2248

Query Match             0.6%;  Score 21;  DB 4;  Length 747;
Best Local Similarity  100.0%;  Pred. No. 5.3;
Matches   21;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy       3248 ATTTTTTATTTTTTAATAA 3268
              ||||||||||||||||||
Db        175 ATTTTTTATTTTTTAATAA 155

RESULT 10
US-08-075-533-1/c
; Sequence 1, Application US/08075533
; Patent No. 5530186
; GENERAL INFORMATION:
;   APPLICANT: Hitz, William D.
;   APPLICANT: Yadav, Narendra S.
;   TITLE OF INVENTION: Nucleotide Sequences of Soybean Acyl-ACP
;   TITLE OF INVENTION: Thioesterase Genes
;   NUMBER OF SEQUENCES: 22
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: E. I. du Pont de Nemours and Company
;     STREET: 1007 Market Street
;     CITY: Wilmington
;     STATE: Delaware
;     COUNTRY: U.S.A
;     ZIP: 19898
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/075,533
;     FILING DATE:
;     CLASSIFICATION: 800
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/631,264
;     FILING DATE: 20-DEC-1990
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Morrissey, Bruce W.
;     REGISTRATION NUMBER: 30,663
;     REFERENCE/DOCKET NUMBER: CR-8926-A
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (302) 992-4927
;     TELEFAX: (302) 892-7949
;     TELEX: 835420
```

```
INFORMATION FOR SEQ ID NO: 1:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1602 base pairs
    TYPE: Nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: cDNA to mRNA
  HYPOTHETICAL: NO
  ANTI-SENSE: NO
  ORIGINAL SOURCE:
    ORGANISM: Glycine max
    STRAIN: Cultivar Wye
    DEVELOPMENTAL STAGE: Early seed fill
    HAPLOTYPE: Diploid
    TISSUE TYPE: Cotyledon
    ORGANELLE: Nucleus
  IMMEDIATE SOURCE:
    LIBRARY: cDNA to mRNA
    CLONE: 22B
  FEATURE:
    NAME/KEY: mat_peptide
    LOCATION: 271..1206
    IDENTIFICATION METHOD: Catalytically active when
    IDENTIFICATION METHOD: expressed in E. coli
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 106..1209
US-08-075-533-1

Query Match          0.6%;   Score 21;  DB 1;  Length 1602;
Best Local Similarity 100.0%;  Pred. No. 5.4;
Matches   21;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   3251 TTTTTATTTTTTTAATAAAAA 3271
          ||||||||||||||||||||
Db     87 TTTTTATTTTTTTAATAAAAA 67

RESULT 11
US-08-948-176-1/c
; Sequence 1, Application US/08948176
; Patent No. 5945585
  GENERAL INFORMATION:
    APPLICANT: HITZ, WILLIAM D.
    APPLICANT: YADAV, NARENDRA S.
    TITLE OF INVENTION: ACYL-ACP THIOESTERASES GENES
    TITLE OF INVENTION: AND THEIR USE IN ALTERING PLANT
    TITLE OF INVENTION: OIL COMPOSITION
    NUMBER OF SEQUENCES: 26
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
      STREET: 1007 MARKET STREET
      CITY: WILMINGTON
      STATE: DELAWARE
      COUNTRY: UNITED STATES OF AMERICA
      ZIP: 19898
    COMPUTER READABLE FORM:
      MEDIUM TYPE: DISKETTE, 3.50 INCH
      COMPUTER: IBM PC COMPATIBLE
      OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
      SOFTWARE: MICROSOFT WORD VERSION 7.0
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/948,176
      FILING DATE:
      CLASSIFICATION: 800
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: 07/631,264
      FILING DATE: DECEMBER 20, 1990
    ATTORNEY/AGENT INFORMATION:
      NAME: CHRISTENBURY, LYNNE M.
      REGISTRATION NUMBER: 30,971
      REFERENCE/DOCKET NUMBER: CR-8926-C
    TELECOMMUNICATION INFORMATION:
```

```
      TELEPHONE: 302-992-5481
      TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 1:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1602 base pairs
    TYPE: Nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: cDNA to mRNA
  HYPOTHETICAL: NO
  ANTI-SENSE: NO
  ORIGINAL SOURCE:
    ORGANISM: Glycine max
    STRAIN: Cultivar Wye
    DEVELOPMENTAL STAGE: Early seed fill
    HAPLOTYPE: Diploid
    TISSUE TYPE: Cotyledon
    ORGANELLE: Nucleus
  IMMEDIATE SOURCE:
    LIBRARY: cDNA to mRNA
    CLONE: 22B
  FEATURE:
    NAME/KEY: mat_peptide
    LOCATION: 271..1206
    IDENTIFICATION METHOD: Catalytically active
    IDENTIFICATION METHOD: when expressed in
    IDENTIFICATION METHOD: E. coli
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 106..1209
US-08-948-176-1

Query Match          0.6%;   Score 21;  DB 2;  Length 1602;
Best Local Similarity 100.0%;  Pred. No. 5.4;
Matches   21;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   3251 TTTTTATTTTTTTAATAAAAA 3271
          ||||||||||||||||||||
Db     87 TTTTTATTTTTTTAATAAAAA 67

RESULT 12
PCT-US91-09160-1/c
; Sequence 1, Application PC/TUS9109160
  GENERAL INFORMATION:
    APPLICANT: Hitz, William D.
    APPLICANT: Yadav, Narendra S.
    TITLE OF INVENTION: Nucleotide Sequences of Soybean Acyl-ACP
    TITLE OF INVENTION: Thioesterase Genes
    NUMBER OF SEQUENCES: 22
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: E. I. du Pont de Nemours and Company
      STREET: 1007 Market Street
      CITY: Wilmington
      STATE: Delaware
      COUNTRY: U.S.A
      ZIP: 19898
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER: PCT/US91/09160
      FILING DATE: 19911216
      CLASSIFICATION:
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/631,264
      FILING DATE: 20-DEC-1990
    ATTORNEY/AGENT INFORMATION:
      NAME: Morrissey, Bruce W.
      REGISTRATION NUMBER: 30,663
      REFERENCE/DOCKET NUMBER: CR-8926-A
```

```
;      TELECOMMUNICATION INFORMATION:
;         TELEPHONE:  (302) 992-4927
;         TELEFAX:  (302) 892-7949
;         TELEX:  835420
;  INFORMATION FOR SEQ ID NO: 1:
;      SEQUENCE CHARACTERISTICS:
;         LENGTH:  1602 base pairs
;         TYPE:  NUCLEIC ACID
;         STRANDEDNESS:  single
;         TOPOLOGY:  linear
;      MOLECULE TYPE:  cDNA to mRNA
;      HYPOTHETICAL:  NO
;      ANTI-SENSE:  NO
;      ORIGINAL SOURCE:
;         ORGANISM:  Glycine max
;         STRAIN:  Cultivar Wye
;         DEVELOPMENTAL STAGE:  Early seed fill
;         HAPLOTYPE:  Diploid
;         TISSUE TYPE:  Cotyledon
;         ORGANELLE:  Nucleus
;      IMMEDIATE SOURCE:
;         LIBRARY:  cDNA to mRNA
;         CLONE:  22B
;      FEATURE:
;         NAME/KEY:  mat_peptide
;         LOCATION:  271..1206
;         IDENTIFICATION METHOD:  Catalytically active when
;         IDENTIFICATION METHOD:  expressed in E. coli
;      FEATURE:
;         NAME/KEY:  CDS
;         LOCATION:  106..1209
PCT-US91-09160-1

Query Match             0.6%;   Score 21;   DB 5;   Length 1602;
Best Local Similarity  100.0%;   Pred. No. 5.4;
Matches    21;   Conservative    0;   Mismatches    0;   Indels    0;   Gaps    0;

Qy     3251 TTTTTATTTTTTAATAAAAA 3271
            |||||||||||||||||||
Db       87 TTTTTATTTTTTAATAAAAA 67

RESULT 13
US-10-204-708-80/c
;  Sequence 80, Application US/10204708
;  Patent No. 6677731
;  GENERAL INFORMATION:
;      APPLICANT:  OLEK, Alexander
;      APPLICANT:  PIEPENBROCK, Christian
;      APPLICANT:  BERLIN, Kurt
;      TITLE OF INVENTION:  Diagnosis of Diseases Associated with DNA Replication
;      TITLE OF INVENTION:  by Assessing DNA Methylation
;      FILE REFERENCE:  5013.1012
;      CURRENT APPLICATION NUMBER:  US/10/204,708
;      CURRENT FILING DATE:  2003-05-06
;      PRIOR APPLICATION NUMBER:  PCT/EP01/03971
;      PRIOR FILING DATE:  2001-04-06
;      PRIOR APPLICATION NUMBER:  DE 10019058.8
;      PRIOR FILING DATE:  2000-04-06
;      PRIOR APPLICATION NUMBER:  DE 10019173.8
;      PRIOR FILING DATE:  2000-04-07
;      PRIOR APPLICATION NUMBER:  DE 10032529.7
;      PRIOR FILING DATE:  2000-06-30
;      PRIOR APPLICATION NUMBER:  DE 10043826.1
;      PRIOR FILING DATE:  2000-09-01
;      NUMBER OF SEQ ID NOS: 98
;  SEQ ID NO 80
;      LENGTH:  8961
;      TYPE:  DNA
;      ORGANISM:  Artificial Sequence
;      FEATURE:
;      OTHER INFORMATION:  chemically treated genomic DNA (Homo sapiens)
;      FEATURE:
```

```
;         NAME/KEY:  unsure
;         LOCATION:  (5096)
;         OTHER INFORMATION:  n is a or g or c or t
US-10-204-708-80

Query Match             0.6%;   Score 21;   DB 4;   Length 8961;
Best Local Similarity  100.0%;   Pred. No. 5.5;
Matches    21;   Conservative    0;   Mismatches    0;   Indels    0;   Gaps    0;

Qy     3251 TTTTTATTTTTTAATAAAAA 3271
            |||||||||||||||||||
Db     1629 TTTTTATTTTTTAATAAAAA 1609

RESULT 14
US-09-134-000C-2872
;  Sequence 2872, Application US/09134000C
;  Patent No. 6617156
;  GENERAL INFORMATION:
;      APPLICANT:  Lynn Doucette-Stamm et al
;      TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;      TITLE OF INVENTION:  ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
;      FILE REFERENCE:  032796-032
;      CURRENT APPLICATION NUMBER:  US/09/134,000C
;      CURRENT FILING DATE:  1998-08-13
;      PRIOR APPLICATION NUMBER:  US 60/055,778
;      PRIOR FILING DATE:  1997-08-15
;      NUMBER OF SEQ ID NOS: 6812
;      SOFTWARE:  PatentIn version 3.1
;  SEQ ID NO 2872
;      LENGTH:  591
;      TYPE:  DNA
;      ORGANISM:  Enterococcus faecalis
US-09-134-000C-2872

Query Match             0.6%;   Score 20;   DB 4;   Length 591;
Best Local Similarity  100.0%;   Pred. No. 16;
Matches    20;   Conservative    0;   Mismatches    0;   Indels    0;   Gaps    0;

Qy     2385 ATCAATTATTATTTTTTTTT 2404
            ||||||||||||||||||||
Db      410 ATCAATTATTATTTTTTTTT 429

RESULT 15
US-08-585-593A-27/c
;  Sequence 27, Application US/08585593A
;  Patent No. 6503706
;  GENERAL INFORMATION:
;      APPLICANT:  ABKEN, Hinrich J
;      APPLICANT:  ALBERT, Winfried
;      APPLICANT:  JUNGFER, Herbert
;      TITLE OF INVENTION:  METHOD OF IDENTIFYING HUMAN AND ANIMAL
;      TITLE OF INVENTION:  CELLS CAPABLE OF UNLIMITED PROLIFERATION OR TUMOR
;      TITLE OF INVENTION:  FORMATION
;      NUMBER OF SEQUENCES:  66
;      CORRESPONDENCE ADDRESS:
;         ADDRESSEE:  Nikaido, Marmelstein, Murray & Oram LLP
;         STREET:  655 Fifteenth Street N.W. Suite 330
;         CITY:  Washington
;         STATE:  D.C.
;         COUNTRY:  U.S.A.
;         ZIP:  20005-5701
;      COMPUTER READABLE FORM:
;         MEDIUM TYPE:  Floppy disk
;         COMPUTER:  IBM PC compatible
;         OPERATING SYSTEM:  PC-DOS/MS-DOS
;         SOFTWARE:  PatentIn Release #1.0, Version #1.30
;      CURRENT APPLICATION DATA:
;         APPLICATION NUMBER:  US/08/585,593A
;         FILING DATE:  16-JAN-1996
;         CLASSIFICATION:  435
;      PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: PCT/EP94/02307
;  FILING DATE: 13-JUL-1994
; PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: DE P 43 23 727.4
;   FILING DATE: 15-JUL-1993
; ATTORNEY/AGENT INFORMATION:
;  NAME: Murray, Robert B.
;  REGISTRATION NUMBER: 22,980
; TELECOMMUNICATION INFORMATION:
;  TELEPHONE: (202)638-5000
;  TELEFAX: (202)638-4810
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
;  LENGTH: 592 base pairs
;  TYPE: nucleic acid
;  STRANDEDNESS: single
;  TOPOLOGY: linear
;  MOLECULE TYPE: DNA
US-08-585-593A-27

Query Match        0.6%;   Score 20;  DB 4;  Length 592;
Best Local Similarity  100.0%;  Pred. No. 16;
Matches  20;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy  3253 TTTATTTTTTTAATAAAAAT 3272
         |||||||||||||||||||||
Db   109 TTTATTTTTTTAATAAAAAT 90

Search completed: July 21, 2004, 19:27:57
Job time : 252 secs
```

```
                OM nucleic - nucleic search, using sw model

                Copyright    GenCore version 5.1.6
                Copyright (c) 1993 - 2004 Compugen Ltd.

Run on:         July 21, 2004, 17:06:46 ; Search time 1433 Seconds
                                           (without alignments)
                                           1616.099 Million cell updates/sec

Title:          US-09-270-437D-6
Perfect score:  3412
Sequence:       1 ggcagcggaggaggcgagga.........aacttgaaatgtttattt 3412

Scoring table:  OLIGO_NUC
                Gapop 60.0 , Gapext 60.0

Searched:       3191023 seqs, 2439312756 residues

Word size :     0

Total number of hits satisfying chosen parameters:        6382046

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :      Published_Applications NA:*
         1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
         2:  /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
         3:  /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
         4:  /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
         5:  /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
         6:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
         7:  /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
         8:  /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
         9:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
        10:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
        11:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
        12:  /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
        13:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
        14:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
        15:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
        16:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
        17:  /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
        18:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
        19:  /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
        19:  /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	3412	100.0	3412	9	US-09-899-651-6	Sequence 6, Appli
2	2953	86.5	3694	9	US-09-764-864-329	Sequence 329, App
3	2143	62.8	3283	9	US-09-899-651-8	Sequence 8, Appli
4	1861	54.5	3905	10	US-09-814-353-20241	Sequence 20241, A
5	1810	53.0	3667	15	US-10-097-340-146	Sequence 146, App
6	1810	53.0	3667	17	US-10-648-593-48	Sequence 48, Appl
7	1496	43.8	2290	9	US-09-764-864-332	Sequence 332, App
8	401	11.8	1186	9	US-09-764-864-749	Sequence 749, App
9	397	11.6	582	10	US-09-814-353-14725	Sequence 14725, A
10	389	11.4	822	9	US-09-764-864-330	Sequence 330, App
11	347	10.2	579	10	US-08-814-353-1995	Sequence 1995, Ap
12	347	10.2	579	10	US-08-814-353-8341	Sequence 8341, Ap
13	333	9.8	472	15	US-10-066-543-2949	Sequence 2949, Ap
14	264	7.7	264	9	US-09-777-564-1124	Sequence 1124, Ap

15	264	7.7	264	15	US-10-015-219-1124	Sequence 1124, Ap
16	263	7.7	263	9	US-09-777-564-1307	Sequence 1307, Ap
17	263	7.7	263	9	US-09-777-564-1537	Sequence 1537, Ap
18	263	7.7	263	15	US-10-015-219-1307	Sequence 1307, Ap
19	263	7.7	263	15	US-10-015-219-1537	Sequence 1537, Ap
20	222	6.5	222	9	US-09-815-34-676	Sequence 676, App
21	222	6.5	222	9	US-09-815-343-1285	Sequence 1285, App
22	222	6.5	222	13	US-10-097-105-676	Sequence 676, App
23	222	6.5	222	13	US-10-097-105-1285	Sequence 1285, App
24	184	5.4	531	15	US-10-066-543-1888	Sequence 1888, Ap
25	167	4.9	218	9	US-09-815-343-1415	Sequence 1415, Ap
26	167	4.9	218	13	US-10-097-105-1415	Sequence 1415, Ap
27	165	4.8	500	9	US-09-783-590-1246	Sequence 1246, Ap
28	143	4.2	402	10	US-09-918-995-5484	Sequence 5484, Ap
29	140	4.1	243	9	US-09-815-343-614	Sequence 614, App
30	140	4.1	243	13	US-10-097-105-614	Sequence 614, App
31	135	4.0	1707	16	US-10-262-445-39	Sequence 39, Appl
32	93	2.7	253	9	US-09-815-343-729	Sequence 729, App
33	93	2.7	253	13	US-10-097-105-729	Sequence 729, App
34	91	2.7	402	9	US-09-783-590-5490	Sequence 5490, Ap
35	83	2.4	380	9	US-09-815-343-349	Sequence 349, App
36	83	2.4	380	9	US-09-880-107-1135	Sequence 1135, Ap
37	81	2.4	590	13	US-10-240-425-760	Sequence 760, App
38	81	2.0	161484	17	US-10-317-401-4	Sequence 4, Appli
39	68	2.0	412	9	US-09-962-436-95	Sequence 95, Appl
40	62	1.8	145	9	US-09-764-864-747	Sequence 747, App
41	55	1.6	399	15	US-10-029-386-14468	Sequence 14468, A
42	55	1.6	567	15	US-10-029-386-765	Sequence 765, App
43	50	1.5	154	9	US-09-783-590-5140	Sequence 5140, Ap
44	37	1.1	495	10	US-09-918-995-26330	Sequence 26330, A
45	26	0.8	26	16	US-10-262-445-95	Sequence 95, Appl

ALIGNMENTS

```
RESULT 1
US-09-899-651-6
; Sequence 6, Application US/0989651
; Patent No. US20020111470A1
; GENERAL INFORMATION:
;  APPLICANT: Chen, Yao-Tseng
;  APPLICANT: Gure, Ali
;  APPLICANT: Tsang, Solam
;  APPLICANT: Stockert, Elisabeth
;  APPLICANT: Jager, Elke
;  APPLICANT: Knuth, Alexander
;  APPLICANT: Old, Lloyd J.
;  TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
;  TITLE OF INVENTION:   Associated
;  TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
;  FILE REFERENCE: LUD 5538
;  CURRENT APPLICATION NUMBER: US/09/899,651
;  CURRENT FILING DATE: 2001-07-06
;  PRIOR APPLICATION NUMBER: US/09/061,709
;  PRIOR FILING DATE: 1998-04-17
;  NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 6
;  LENGTH: 3412
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
US-09-899-651-6

  Query Match             100.0%;  Score 3412;  DB 9;  Length 3412;
  Best Local Similarity   100.0%;  Pred. No.0;
  Matches 3412;  Conservative   0;  Mismatches 0;  Indels 0;  Gaps  0;

Qy        1 GGCAGCGGAGGAGGCGAGGAGGCGGCGACGCCGGGTACCCGGCGGGGGAGCCGGGCTTCGGCGG 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 GGCAGCGGAGGAGGCGAGGAGGCGGCGACGCCGGGTACCCGGCGGGGGAGCCGGGCTTCGGCGG 60

Qy       61 AAGAGACCGATGATGATGAACAAGCTTTACATCGGGAACTGAGCCCCGGCCGTCACCGCCGAC 120
```

Thu Jul 22 08:21:42 2004

us-09-270-437d-6.Oligo.rnpb

```
Qy  3361 AGCCAAGAACCNATATGGCCTTCTTTTTGGACAAAACCTTGAAAATGTTTATTT 3412
         |||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db  3361 AGCCAAGAACCNATATGGCCTTCTTTTTGGACAAAACCTTGAAAATGTTTATTT 3412
```

RESULT 2
US-09-764-864-329
; Sequence 329, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PF223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 329
; LENGTH: 3694
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-329

Query Match   86.5%;   Score 2953;   DB 9;   Length 3694;
Best Local Similarity 99.9%;   Pred. No. 0;
Matches 3303;   Conservative 0;   Mismatches 0;   Indels 3;   Gaps 2;

```
Qy   68 GGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCCGCCGTCACCCGCGACGACCTCC 127
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   91 GGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCCGCCGTCACCCGCGACGACCTCC 150

Qy  128 GGGCAGCTCTTTGGGGACAGGAGCTGCCCCTGCGGGACAGGTCCTGCTGAAGTCCGGCT 187
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  151 GGGCAGCTCTTTGGGGACAGGAGCTGCCCCTGCGGGACAGGTCCTGCTGAAGTCCGGCT 210

Qy  188 ACGGCCTTCGTGGACTACCCCGACCAGAGAACTGGGCCATCGCGGCCAGCCCTCTCGG 247
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  211 ACGGCCTTCGTGGACTACCCCGACCAGAGAACTGGGCCATCGCGGCCAGCCCTCTCGG 270

Qy  248 GTAAAGTGGAATTGCATGGGAAATCATGGAGAGTTGATTACTCAGTCTCTAAAAGCTAA 307
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  271 GTAAAGTGGAATTGCATGGGAAATCATGGAGAGTTGATTACTCAGTCTCTAAAAGCTAA 330

Qy  308 GGAGCAGGAGAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGAGGTGTTGGATG 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  331 GGAGCAGGAGAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGAGGTGTTGGATG 390

Qy  368 GACTTTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGACAGAGAAA 427
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  391 GACTTTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGACAGAGAAA 450

Qy  428 CCGCCGTGTCAACGTCACATATGCAACAAGAGGAGCAAAAATAGCCATGGAGAAGAGC 487
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  451 CCGCCGTGTCAACGTCACATATGCAACAAGAGGAGCAAAAATAGCCATGGAGAAGAGC 510

Qy  488 TAAGCGGGGCATCAGTTTGAGAACTACTCCTTCAGATTCCTACATCCCGGATGAAGAGG 547
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  511 TAAGCGGGGCATCAGTTTGAGAACTACTCCTTCAGATTCCTACATCCCGGATGAAGAGG 570

Qy  548 TGAGCTCCCCTTCGCCCCCCAGCGTGGGGACCCACTCTTCCCGGGAGCCAAG 607
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  571 TGAGCTCCCCTTCGCCCCCCAGCGTGGGGACCCACTCTTCCCGGGAGCCAAG 630

Qy  608 GCCACGCGCCCTGGGGGCCACTTCCAGGCCAGACAGATTGATTTCCCGCGCCGATCCTGG 667
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  631 GCCACGCGCCCTGGGGGCCACTTCCAGGCCAGACAGATTGATTTCCCGCGCCGATCCTGG 690

Qy  668 TCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAACATCA 727
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  691 TCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAACATCA 750

Qy  728 CTAAGCAGACCCAGTCCCGGTAGATATCCCGGTAGATATCCATAGAAGAGAACTCTGAGCTGCAGAGA 787
```

```
Qy 2281 GGAAGATGTTAAGAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCGCGGTAATATAT 2340
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2281 GGAAGATGTTAAGAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCGCGGTAATATAT 2340

Qy 2341 TTTAGAAATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTTT 2400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2341 TTTAGAAATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTTT 2400

Qy 2401 TTTTTCTTTTAAAGAGAAAGCAGCCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG 2460
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2401 TTTTTCTTTTAAAGAGAAAGCAGCCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG 2460

Qy 2461 GTCTCACGGTGTAGAGAGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGAGGGAAAT 2520
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2461 GTCTCACGGTGTAGAGAGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGAGGGAAAT 2520

Qy 2521 CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAAGGGATA 2580
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2521 CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAAGGGATA 2580

Qy 2581 CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTCATGGCCGTGTCTGCTATTCTCTGTCT 2640
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2581 CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTCATGGCCGTGTCTGCTATTCTCTGTCT 2640

Qy 2641 TCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCT 2700
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2641 TCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCT 2700

Qy 2701 CTCTGTTTATCTCTCCCCCTTCTTCTCCATCTCCATTCTTTTGAA 2760
         |||||||||||||||||||||||||||||||||||||||||||||
Db 2701 CTCTGTTTATCTCTCCCCCTTCTTCTCCATCTCCATTCTTTTGAA 2760

Qy 2761 TTTCCATCCTCCGTATCTCACGCACCCCCCCCCCCAGGCAAAGCA 2820
         |||||||||||||||||||||||||||||||||||||||||||||
Db 2761 TTTCCATCCTCCGTATCTCACGCACCCCCCCCCCCAGGCAAAGCA 2820

Qy 2821 GTGCTCTGAGTATCACATCACACAAAAGAACACAAAGCGAAACACCAGCCTCAA 2880
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2821 GTGCTCTGAGTATCACATCACACAAAAGAACACAAAGCGAAACACCAGCCTCAA 2880

Qy 2881 CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGGAAG 2940
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2881 CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGGAAG 2940

Qy 2941 AGGAAAACAGGAACCCACCACAAAACAACCAACCAATCAACCAAACAAGAAAAAATTCCACAATG 3000
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2941 AGGAAAACAGGAACCCACCACAAAACAACCAACCAATCAACCAAACAAGAAAAAATTCCACAATG 3000

Qy 3001 AAAGAATGTATTTTGTCTTTTGTGTATAAGCCATCAATATTCAGCAAAATG 3060
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3001 AAAGAATGTATTTTGCTTTTTGTGTATAAGCCATCATATTCAGCAAAATG 3060

Qy 3061 ATTCCTTTCTTTAAAAAAAAAATGTGGAGGAAGTAGAAGTAGAAATTTACCAAGGTTGTTGTTGGCC 3120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3061 ATTCCTTTCTTTAAAAAAAAAATGTGGAGGAAGTAGAAGTAGAAATTTACCAAGGTTGTTGTTGGCC 3120

Qy 3121 CAGGGGGTTAAATTCACAGATTTTTTAATAAAATGAGTTGGAAAAATAAGATATCAACT 3180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3121 CAGGGGGTTAAATTCACAGATTTTTTAATAAAATGAGTTGGAAAAATAAGATATCAACT 3180

Qy 3181 GTGTTTTTACCTCAGCCACCTTGCTCTTGTCTTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAG 3240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3181 GTGTTTTTACCTCAGCCACCTTGCTCTTGTCTTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAG 3240

Qy 3241 TTGGAGCATTTTTATTTTTATTTTTAATAAAATGAGTTGGAAAAATAAGATATCAACT 3300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3241 TTGGAGCATTTTTATTTTTATTTTTAATAAAATGAGTTGGAAAAATAAGATATCAACT 3300

Qy 3301 GCCAGCCTGGAGAAGGTGACAGTGCACAGCTGTTCTCCGCT 3360
         |||||||||||||||||||||||||||||||||||||||||
Db 3301 GCCAGCCTGGAGAAGGTGACAGTGCACAGCTGTTCTCCGCT 3360
```

This page contains a DNA sequence alignment (query "Qy" vs database "Db") in monospaced format with position numbers. The alignment rows are paired as follows, with the first (lower-numbered) block on the left portion of the page and the continuation on the right portion:

```
Qy   751  ...CTAAGCAGACCCAGTCCGGGTAGATATCCATAGAAAAGAACTTCGGAGCTCCAGAGA  810
Db  1831  AACAGGTGAAGCAGCAGGAGGAGCAGGAGAAATACCCTCAGGGAGTCGCCTCCACAGGCGCAGCAAGT  1890

Qy   788  AGCCTTGTCACCATCCATGCCAAGGGGACTTCTGAAGCATGCCCATGATTCTTG  847
Db  1868  GAGGCTCCCACAGGCACCAGGCAAAACAACGATGAATGTAGCCCTTCCAACACCTGACAG  1927
Db  1891  GAGGCTCCCACAGGCACCAGGCAAAACAACGATGAATGTAGCCCTTCCAACACCTGACAG  1950

Qy   848  AAATCATGCAGAAAGAGGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAATCT  907
Db  1928  AATGAGACCAAACGCAGCGCCAGATCGGGAGGCAAACCAAAGACCATCTGAGGATGAG  1987
Db  1951  AATGAGACCAAACGCAGCGCCAGATCGGGAGGCAAACCAAAGACCATCTGAGGATGAG  2010

Qy   908  TGGCCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGAGGCAGAAATTTGAAGAAAA  967
Db  1988  AAGTCTGCGGAGGCGGCCCAGGGACTCTGCCGAGGCCCTGAGAACCCGAGGGCCGAGGAG  2047
Db  2011  AAGTCTGCGGAGGCGGCCCAGGGACTCTGCCGAGGCCCTGAGAACCCGAGGGGCGGAGGAG  2070

Qy   968  TTGAACATGAAACAGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATATACA  1027
Db  2048  GGGCGGGGAAGTCAGCCAGTTTGGCCAGACCACCGAGCCCGCCTCCCGCCTCCTCCAGG  2107
Db  2071  GGGCGGGGAAGTCAGCCAGTTTGGCCAGAACACCGAGCCCGCCTCCCGCCTCCCCCAGG  2130

Qy  1028  ACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGAGATAG  1087
Db  2108  GCTTCTGCGAGGCTTCAGCCATCCACTTCACCATCCACTCCTGAACTCCCAC  2167
Db  2131  GCTTCTGCGAGGCTTCAGCCATCCACTTCACCATCCACTCCTGAACTCCCAC  2190

Qy  1088  AGATTATGAAGAAGTGCGTGAGGGCTTTGAAAATGATATGCTGGCTGTTAACCAACAAG  1147
Db  2168  GACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGC  2227
Db  2191  GACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGC  2250

Qy  1148  CCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGACTGTCCG  1207
Db  2228  ACACCCTTTTCTGTCGGCAAATCGTCTCGTGTACATGTGTGTACATATTAGAAAGGGAAGA  2287
Db  2251  ACACCCTTTTCTGTCGGCAAATCGTCTCGTGTACATGTGTGTACATATTAGAAAGGGAAGA  2310

Qy  1208  TGCTATCTCCACCAGCAGGGCCTCCCGCGGAGCTCCCCGTGCCCCCTACCACCCCTTCA  1267
Db  2288  TGTTAAGATATGTGGCCGTGGGTTACACAGGGTGCCTGCGAGCGGTAATATTTAGAA  2347
Db  2311  TGTTAAGATATGTGGCCGTGGGTTACACAGGGTGCCTGCGAGCGGTAATATATTTAGAA  2370

Qy  1268  CTACCCACTCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTTGGCCCCGTTCCCGC  1327
Db  2348  ATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTATTTTTTTCT  2407
Db  2371  ATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTATTTTTTTCT  2430

Qy  1328  ATCATCACTCTTATCCAGACAGGAGATTGTGAATCTTCATCCCAAACCAGGCTGTGTGG  1387
Db  2408  TTTTAAGAGAAAGCAGGCTTTCTAGACTTTAAAGAATAAAGTCTTGGGAGGTCTCAC  2467
Db  2431  TTTTAAGAGAAAGCAGGCTTTCTAGACTTTAAAGAATAAAGTCTTGGGAGGTCTCAC  2490

Qy  1388  GCGCCATCATCGGGAAGAAGGGGGCACACATCAAACAGCTGGCGAAAGGCCT  1447
Db  2468  GGTGTAGAGAGGAGCTTTGAGGCCACCCGACAAAATTCACCCAGAGGGAAATCTCGTCG  2527
Db  2491  GGTGTAGAGAGGAGCTTTGAGGCCACCCGACAAAATTCACCCAGAGGGAAATCTCGTCG  2550

Qy  1448  CTATCAAGATTGCCCCTGCGGAAGGCCAGCGAAGGATGGTCATCATCCACCG  1507
Db  2528  GAAGGACACTCACGGCCAGTTCTGGATCACCTGTGTATGTCAACAGAGGGATACCGTCTC  2587
Db  2551  GAAGGACACTCACGGCCAGTTCTGGATCACCTGTGTATGTCAACAGAGGGATACCGTCTC  2610

Qy  1508  GGGCACCCGAAGCCCAGTTCAAGCCCAGGGACGATCTTTGGGAAACTGAAAGAGAGAA  1567
Db  2588  CTTGAAGAGGAAACTCGTCTCACTCCTCATGCCTGTCTAGCTCATCACCCATTTCTCTTT  2647
Db  2611  CTTGAAGAGGAAACTCGTCTCACTCCTCATGCCTGTCTAGCTCATCACCCATTTCTCTTT  2670

Qy  1568  ACTTCTTTAACCCAAAGAGAAGTGAAGCTGAAGCGCATATCAGATGCCCTCTTCCA  1627
Db  2648  GCTTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCTCTCTG  2707
Db  2671  GCTTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCTCTCTG  2730

Qy  1628  CAGCTGGCCGGGTGATTGGCAAAGGTGGCAAGACCGTTGCAAGCGCCAAGAACTGTCAGAACTGAATGTTAACCA  1687
Db  2708  TTTATCTCTCCCCCTCCCCCTCCCCCTTCTTCTCCATTCTTTGAATTCCTC  2767
Db  2731  TTTATCTCTCCCCCTCCCCCTCCCCCTTCTTCTCCATTCTTTGAATTCCTC  2790

Qy  1688  GTGCAGAGTCATCGTGCCTGTGACCAAACGCCAGATGAAAATGAAGTGATCGTCA  1747
Db  2768  ATCCCTCCATCTCAATCCGTATCTCACGA--CCCCCCCCCCCCCCCAGCAAAGCAGTGCT  2825
Db  2791  ATCCCTCCATCTCAATCCGTATCTCACGCACCCCCCCCCCCCCCAGCAAAGCAGTGCT  2850

Qy  1748  GAATTATCGGGCACTTCTTTGCTGACCAGCCGGAAATTGTAC  1807
Db  2826  CTGAGTATCACATCACACAAAAGAACAAAAGCGAAACACAAACCAGCCTCAACTTAC  2885
Db  2851  CTGAGTATCACATCACACAAAAGAACAAAAGCGAAACACAAACCAGCCTCAACTTAC  2910

Qy  1808  AACAGGTGAAGCAGCAGGAGGAGCAGGAGAAATACCCTCAGGGAGTCGCCTCCACAGGCGCAGCAAGT  1867
Db  2886  ACTTGGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGGAA  2945
Db  2911  ACTTGGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGGAA  2970
```

```
Qy  1390 GCCATCATCGGGAAGAAGGAGGGGGGCCACACATCAAATCAACAGCTGGCCGAGATTCGCCGCGAGCTCT 1449
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 GCCATCATCGGGAAGAAGGAGGGGGGCCACACATCAAATCAACAGCTGGCCGAGATTCGCCGCGAGCTCT 1320

Qy  1450 ATCAAGATTGCCCCTGCGGAAGGCCCAGACGTTCAGCGAAAGATGGTCATCATCACCGGG 1509
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 ATCAAGATTGCCCCTGCGGAAGGCCCAGACGTTCAGCGAAAGATGGTCATCATCACCGGG 1380

Qy  1510 CCACCGGAAGCCCAGTTCAAGGCCCAGGTCAAGGCCAGGACGGATCTTTGGGAAACTGAAGAGGGAAAAC 1569
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 CCACCGGAAGCCCAGTTCAAGGCCCAGGTCAAGGCCAGGACGGATCTTTGGGAAACTGAAGAGGGAAAAC 1440

Qy  1570 TTCTTTAACCCCAAAGAAGAAGAGTGAAGCGCCATATCAGAGGTGCCCCTTCCACA 1629
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 TTCTTTAACCCCAAAGAAGAAGAGTGAAGCGCCATATCAGAGGTGCCCCTTCCACA 1500

Qy  1630 GCTGGCCGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTGCAGAACTTAACCAGT 1689
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 GCTGGCCGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTGCAGAACTTAACCAGT 1560

Qy  1690 GCAGAAGTCATCGTGCCTCGTGACCAAAGCCAGATGAAAATGAGGAAGTGATCGTCAGA 1749
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 GCAGAAGTCATCGTGCCTCGTGACCAAAGCCAGATGAAAATGAGGAAGTGATCGTCAGA 1620

Qy  1750 ATTATCGGGCACTTCTTTGCTAGCCAGAGCTGTTCTGAATTGTCTTCCGGTACAA 1809
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621 ATTATCGGGCACTTCTTTGCTAGCCAGAGCTGTTCTGAATTGTCTTCCGGTACAA 1680

Qy  1810 CAGGTGAAGCAGCAGCACCAGCAGCAAAATACCCTCAGGGAGTGCGCTCACAGCGCAGCAAGTGA 1869
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681 CAGGTGAAGCAGCAGCACCAGCAGCAAAATACCCTCAGGGAGTGCGCTCACAGCGCAGCAAGTGA 1740

Qy  1870 GGCTCCCACAGGCACCAGCAGCAAAACAACGGATGAATGAATGAGCCTTCCAACACTGACAGAA 1929
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741 GGCTCCCACAGGCACCAGCAGCAAAACAACGGATGAATGAATGAGCCTTCCAACACTGACAGAA 1800

Qy  1930 TGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAAGACCATCTGAGGAATGAGAA 1989
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801 TGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAAGACCATCTGAGGAATGAGAA 1860

Qy  1990 GTCTGCGGAGGCCGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGCCCAGGAGGG 2049
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861 GTCTGCGGAGGCCGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGCCCAGGAGGG 1920

Qy  2050 GCGGGGAAGGTCAGCCAGGTTGCAGGCCAGGTTGCCAGAACCACCAGCCCCGCTCCCGGCCCTCCCGCCCTCCCCCAGGGC 2109
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921 GCGGGGAAGGTCAGCCAGGTTGCAGGCCAGGTTGCCAGAACCACCAGCCCCGCTCCCGGCCCTCCCGCCCTCCCCCAGGGC 1980

Qy  2110 TTCTGCAGGCTTCAGCCATCCACTTCACCATCCACTCGGATCTCTCCTGAACTCCCAGA 2169
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981 TTCTGCAGGCTTCAGCCATCCACTTCACCATCCACTCGGATCTCTCCTCCCACCA 2040

Qy  2170 CGTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGCAC 2229
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041 CGTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGCAC 2100

Qy  2230 ACCCTTTTCTGTGGCAAATCGTCTCTGCAATCGTGTGTACATATTAGAAAGGGAAGATG 2289
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101 ACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATATTAGAAAGGGAAGATG 2160

Qy  2290 TTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCACAGGGTAATATATTTTAGAAAT 2349
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161 TTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCACAGGGTAATATATTTTAGAAAT 2220

Qy  2350 AATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTATTTTTTCTTT 2409
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221 AATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTATTTTTCTTT 2280

Qy  2410 TTAAAGAGAAAGCAGCTTTTCAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTGTGGGC 2469
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281 TTAAAGAGAAAGCAGCTTTTCAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTGTGGGC 2340
```

```
Qy  2946 AACAGGAACCCACCAAACCAACCAACCAATCAACCAAACAAAGAAAAATTCCACAATGAAAGA 3005
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2971 AACAGGAACCCACCAAACCAACCAACCAATCAACCAAACAAAGAAAAATTCCACAATGAAAGA 3030

Qy  3006 ATGATATTTGTCTTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCC 3065
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3031 ATGATATTTGTCTTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCC 3090

Qy  3066 TTTCTTT-AAAAAAAAAATGTGGAGAAAGTAGAAATTTACCAAGGTTGTTGGCCCAGG 3124
         ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3091 TTTCTTTAAAAAAAAAATGTGGAGAAAGTAGAAATTTACCAAGGTTGTTGGCCCCAGG 3150

Qy  3125 GCGTTAAATTCACAGATTTTTTAACGAGAAAACACACAGAGAAGAAGCTACCTCAGTGT 3184
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3151 GCGTTAAATTCACAGATTTTTTAACGAGAAAACACACAGAGAAGAAGCTACCTCAGTGT 3210

Qy  3185 TTTTACCTCAGCACCTTGCTCTTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGG 3244
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3211 TTTTACCTCAGCACCTTGCTCTTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGG 3270

Qy  3245 AGCATTTTTTTATTTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATATCAACTGCCA 3304
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3271 AGCATTTTTTTATTTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACTGCCA 3330

Qy  3305 GCCTGGAGAAGGTGACAGTCCAAGTGCTGCGACAGCTGTTCTGAATTGTCTTCCGCTAGCC 3364
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3331 GCCTGGAGAAGGTGACAGTCCAAGTGCTGCGACAGCTGTTCTGAATTGTCTTCCGCTAGCC 3390

Qy  3365 AAGAACC 3371
         |||||||
Db  3391 AAGAACC 3397
```

```
RESULT 3
US-09-899-651-8
; Sequence 8, Application US/09899651
; Patent No. US20020111470A1
; GENERAL INFORMATION:
;  APPLICANT: Chen, Yao-Tseng
;  APPLICANT: Gure, Ali
;  APPLICANT: Tsang, Solam
;  APPLICANT: Stockert, Elisabeth
;  APPLICANT: Jager, Elke
;  APPLICANT: Knuth, Alexander
;  APPLICANT: Old, Lloyd J.
;  TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
;  TITLE OF INVENTION: Associated
;  TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
;  FILE REFERENCE: LUD 5538
;  CURRENT APPLICATION NUMBER: US/09/899,651
;  CURRENT FILING DATE: 2001-07-06
;  PRIOR APPLICATION NUMBER: US/09/061,709
;  PRIOR FILING DATE: 1998-04-17
;  NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 8
;  LENGTH: 3283
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
US-09-899-651-8

  Query Match        62.8%;  Score 2143;  DB 9;  Length 3283;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2143;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy  1270 ACCCACTCCGGATACTTCTCCAGCAGGAGATTGTCACCCCATCACCAGGTTGGCCCGTTCCCGCAT 1329
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 ACCCACTCCGGATACTTCTCCAGCAGGAGATTGTCACCCCATCACCAGGTTGGCCCGTTCCCGCAT 1200

Qy  1330 CATCACTCTTATCCAGCAGGAGATTGTGAATCTTCCTTCATCCCAACCCAGGCTGTGGGGC 1389
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 CATCACTCTTATCCAGCAGGAGATTGTGAATCTTCCTTCATCCCAACCCAGGCTGTGGGGC 1260
```

```
    TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
    TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
    TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
    FILE REFERENCE: MRI-006B
    CURRENT APPLICATION NUMBER: US/09/814,353
    CURRENT FILING DATE: 2001-03-21
    PRIOR APPLICATION NUMBER: US 60/191,031
    PRIOR FILING DATE: 2000-03-21
    PRIOR APPLICATION NUMBER: US 60/207,124
    PRIOR FILING DATE: 2000-05-25
    PRIOR APPLICATION NUMBER: US 60/211,940
    PRIOR FILING DATE: 2000-06-15
    PRIOR APPLICATION NUMBER: US 60/216,820
    PRIOR FILING DATE: 2000-07-07
    PRIOR APPLICATION NUMBER: US 60/220,661
    PRIOR FILING DATE: 2000-07-25
    PRIOR APPLICATION NUMBER: US 60/257,672
    PRIOR FILING DATE: 2000-12-21
    NUMBER OF SEQ ID NOS: 22037
    SOFTWARE: FastSEQ for Windows Version 4.0

    SEQ ID NO 20241
    LENGTH: 3905
    TYPE: DNA
    ORGANISM: Homo sapiens
    FEATURE:
    NAME/KEY: misc_feature
    LOCATION: 1, 3897, 3898, 3899, 3900, 3901, 3902, 3903, 3904, 3905
    OTHER INFORMATION: n = A,T,C or G
US-09-814-353-20241

  Query Match             54.5%;  Score 1861;  DB 10;  Length 3905;
  Best Local Similarity   99.9%;
  Matches 2101;  Mismatches    0;  Conservative    0;  Indels    2;  Gaps    2;
```

```
Qy   2470 TGTAGAGGAGGCTTTGAGGCCACCCGCACAAAATTCACCCAGAGGGAAATCTCGTCGGA 2529
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 TGTAGAGGAGGCTTTGAGGCCACCCGCACAAAATTCACCCAGAGGGAAATCTCGTCGGA 2400

Qy   1270 ACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGCCCGTTCCCGCAT 1329
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1507 ACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGCCCGTTCCCGCAT 1566

Qy   2530 AGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAGAGGATACCGTCTCCT 2589
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 AGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAGAGGATACCGTCTCCT 2460

Qy   1330 CATCACTCTTATCCAGACAGGAGATTGTGAATCTTCATCCCAACCCAGGCTGTGTGGGC 1389
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1567 CATCACTCTTATCCAGACAGGAGATTGTGAATCTTCATCCCAACCCAGGCTGTGTGGGC 1626

Qy   2590 TGAAGAGGAAACTCTGTCACTCCTCAGCTCATACCCCATTTCTCTTTGC 2649
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 TGAAGAGGAAACTCTGTCACTCCTCAGCTCATACCCCATTTCTCTTTGC 2520

Qy   1390 GCCATCATCGGGAAGAAGGGGGCACACATCAAACAGCTGGCGAGATTCGGCGGAGCCTCT 1449
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1627 GCCATCATCGGGAAGAAGGGGGCACACATCAAACAGCTGGCGAGATTCGGCGGAGCCTCT 1686

Qy   2650 TTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCTCTCTGTT 2709
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 TTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCTCTCTGTT 2580

Qy   1450 ATCAAGATTGCCCCTGCGGAAGGCCCAGAGCGTCAGCGGAAAGGATGTCATCATCACCGGG 1509
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1687 ATCAAGATTGCCCCTGCGGAAGGCCCAGAGCGTCAGCGGAAAGGATGTCATCATCACCGGG 1746

Qy   2710 TATCTCTCCCCTCCCCCCCTCCCTTCTCCATCCCATCTCTTTGAATTCCTCAT 2769
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 TATCTCTCCCCTCCCCCCCTCCCTTCTCCATCCCATCTCTTTGAATTCCTCAT 2640

Qy   1510 CCACCGGAAGCCCAGTTCAAGGCCCCAGGGGACGGGATCTTTGGGAAACTGAAAACTGAAAAC 1569
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1747 CCACCGGAAGCCCAGTTCAAGGCCCCAGGGGACGGGATCTTTGGGAAACTGAAAACTGAAAAC 1806

Qy   2770 CCCTCCCATCTCAATCCCGTATCTACGCACCCCCCCCCCCCCCCCCAGGCGAAAGCAGTGTCTGA 2829
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 CCCTCCCATCTCAATCCCGTATCTACGCACCCCCCCCCCCCCCCCCCAGGCGAAAGCAGTGTCTGA 2700

Qy   1570 TTCTTTAACCCAAAGAAGAAGTGAAGCTGAACGGCCAATATCAGAGTGCCCTCCTTCCACA 1629
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1807 TTCTTTAACCCAAAGAAGAAGTGAAGCTGAACGGCCAATATCAGAGTGCCCTCCTTCCACA 1866

Qy   2830 GTATCACATCACAAGAAGGCGAAACACACAAAACCAGCCTCAACTTACACTT 2889
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 GTATCACATCACAAGAAGGCGAAACACACAAAACCAGCCTCAACTTACACTT 2760

Qy   1630 GCTGGCCGGGTGATTGGCAAGAGTGGCAAGACCGTGAACGACCGTGAACGACCTTAACCAGT 1689
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1867 GCTGGCCGGGTGATTGGCAAGAGTGGCAAGACCGTGAACGACCGTGAACGACCTTAACCAGT 1926

Qy   2890 GGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGGAAAACA 2949
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 GGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGGAAAACA 2820

Qy   1690 GCAGAAGTCATCGTGCCTCGTTGACCAAACGCCAGATGAAAATGAAAATGAGGAAGTGATCGTCAGA 1749
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1927 GCAGAAGTCATCGTGCCTCGTTGACCAAACGCCAGATGAAAATGAAAATGAGGAAGTGATCGTCAGA 1986

Qy   2950 GGAACCCCACCAAAACCAATCAACCAAACAAAGAAAAAATTCCACAATGAAAGAATGT 3009
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 GGAACCCCACCAAAACCAATCAACCAAACAAAGAAAAAATTCCACAATGAAAGAATGT 2880

Qy   1750 ATTATCGGGACACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCGCAGGGAAATTGTACAA 1809
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1987 ATTATCGGGACACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCGCAGGGAAATTGTACAA 2046

Qy   3010 ATTTTGTCTTTTTGCATTTGGTGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTC 3069
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 ATTTTGTCTTTTTGCATTTGGTGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTC 2940

Qy   1810 CAGGTGAAGCAGCAGGAGCAGGAGCAGAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAGCAAGTGA 1869
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2047 CAGGTGAAGCAGCAGCAGGAGCAGGAGCAGAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAGCAAGTGA 2106

Qy   3070 TTTTTAATTTTTTTAATAAAAATGAGTTGGAAAAATAAGATAATCAACTGCCAGCCTG 3129
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 TTTTTAATTTTTTTAATAAAAATGAGTTGGAAAAATAAGATAATCAACTGCCAGCCTG 3000

Qy   3130 AAATTCACAGATTTTTTAACGAGAAAACACACAGAAGAAGCTACTTCAGTGTGTTTTA 3189
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 AAATTCACAGATTTTTTTAACGAGAAAACACACAGAAGAAGCTACTTCAGTGTGTTTTA 3060

Qy   3190 CCTCAGCACCTTGCTCTGTGTTTCCCTTAGAGTTTGTAAAGCTGATAGTTGGAGCAT 3249
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 CCTCAGCACCTTGCTCTGTGTTTCCCTTAGAGTTTGTAAAGCTGATAGTTGGAGCAT 3120

Qy   3250 TTTTTTATTTTTTAATAAAAATGAGTTGGAAAAATAAGATAATCAACTGCCAGCCTG 3309
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 TTTTTTATTTTTTTAATAAAAATGAGTTGGAAAAATAAGATAATCAACTGCCAGCCTG 3180

Qy   3310 GAGAAGGTGACAGTGCAACAGTGTTGCAACAGCTGTTCTTCCGCTAGCCAAGAA 3369
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 GAGAAGGTGACAGTGCCAACAGTGTTGCAACAGCTGTTCTTCCGCTAGCCAAGAA 3240

Qy   3370 CCNATATGCCCTTCTTTTTGGACAAACCTTGAAAATGTTTATTT 3412
          |||||||||||||||||||||||||||||||||||||||||||
Db   3241 CCNATATGCCCTTCTTTTTGGACAAACCTTGAAAATGTTTATTT 3283
```

```
RESULT 4
US-09-814-353-20241
; Sequence 20241, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
    APPLICANT: Lee, John
    APPLICANT: Thompson, Pamela
    APPLICANT: Lillie, James
```

```
Qy  1870 GGCTCCCACAGGCACCAGCAAAACAACGATGAATGTAGCCCTTCCAACACCTGACAGAA 1929
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2107 GGCTCCCACAGGCACCAGCAAAACAACGATGAATGTAGCCCTTCCAACACCTGACAGAA 2166

Qy  1930 TGAGACCAAACGCAGCCAGCCAGCCAGATCGGAGCAAACCAAAGACCATCTGAGGAATGAGAA 1989
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2167 TGAGACCAAACGCAGCCAGCCAGCCAGATCGGAGCAAACCAAAGACCATCTGAGGAATGAGAA 2226

Qy  1990 GTCTGCGGAGGAGGCGGCCAGGGACTCTGCCGCCAGGGACCCTGAGAACCCCAGGGGCCGAGGAGGGG 2049
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2227 GTCTGCGGAGGAGGCGGCCAGGGACTCTGCCGCCAGGGACCCTGAGAACCCCAGGGGCCGAGGAGGGG 2286

Qy  2050 GCGGGGAAGGTCAGCCAGGTTGCCAGAACCACCGAGCCCCGCCTCCCGCCTCCCGCCCAGGGC 2109
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2287 GCGGGGAAGGTCAGCCAGGTTGCCAGAACCACCGAGCCCCGCCTCCCGCCTCCCGCCCAGGGC 2346

Qy  2110 TTCTGCAGGCTTCAGCCATCCACTTCACCATCCGATCTCTCCTGAACTCCCACGA 2169
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2347 TTCTGCAGGCTTCAGCCATCCACTTCACCATCCGATCTCTCCTGAACTCCCACGA 2406

Qy  2170 CGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGCAC 2229
         |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db  2407 CGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGCAC 2466

Qy  2230 ACCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATATTAGAAAGGGAAGATG 2289
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2467 ACCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATATTAGAAAGGGAAGATG 2526

Qy  2290 TTAAGATATGTGGCCTGTGGTTACACAGGGTGCCTGCAGCGGGTAATAATATATTTAGAAAT 2349
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2527 TTAAGATATGTGGCCTGTGGTTACACAGGGTGCCTGCAGCGGGTAATAATATATTTAGAAAT 2586

Qy  2350 AATATATCAAATAACTAACTAACTAACTTCCAATTTTAATCAATTATTAATTTTTTTTTTCTTT 2409
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2587 AATATATCAAATAACTAACTAACTAACTTCCAATTTTAATCAATTATTAATTTTTTTTTTCTTT 2646

Qy  2410 TTAAAGAGAAAGCAGCGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTCACGG 2469
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2647 TTAAAGAGAAAGCAGCGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTCACGG 2706

Qy  2470 TGTGAGAGAGGAGCTTTGAGGCCACCCGACACAAATTCACCCAGAGGGAAATCTCGTCGGA 2529
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2707 TGTGAGAGAGGAGCTTTGAGGCCACCCGACACAAATTCACCCAGAGGGAAATCTCGTCGGA 2766

Qy  2530 AGGACACTCACGGCAGTTCTCGGATCACCTGTATGTCAACAGAAGGGATACCGTCTCCT 2589
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2767 AGGACACTCACGGCAGTTCTCGGATCACCTGTATGTCAACAGAAGGGATACCGTCTCCT 2826

Qy  2590 TGAAGAGGAAACTCTGCACTCCTATGCCTGTCCAGCTCATACACCCATTTCTCTTTGC 2649
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2827 TGAAGAGGAAACTCTGCACTCCTATGCCTGTCCAGCTCATACACCCATTTCTCTTTGC 2886

Qy  2650 TTCACAGGTTTTTAAACTGGTTTTTTGCATATGCTATATATAATTCTCTGTCTGTCTGTT 2709
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2887 TTCACAGGTTTTTAAACTGGTTTTTTGCATATGCTATATATAATTCTCTGTCTGTCTGTT 2946

Qy  2710 TATCTCTCCCCTCCCCTCCCCTCCCCTTCTTCTCTCCATCTCCATTCTTTTGAATTCCTCAT 2769
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2947 TATCTCTCCCCTCCCCTCCCCTCCCCTTCTTCTCTCCATCTCCATTCTTTTGAATTCCTCAT 3006

Qy  2770 CCCTCCATCTCAATCCCGTATCTACGGCACCCCCCCCCCCCCCCCAGGCAAAGCAGTGCTCTGA 2829
         ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db  3007 CCCTCCATCTCAATCCCGTATCTACGGCA-CCCCCCCCCCCCCCCAGGCAAAGCAGTGCTCTGA 3065

Qy  2830 GTATCACATCACACAAAAGGAACAAAAGCGAAACCAGGCTCAACTTACACTT 2889
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3066 GTATCACATCACACAAAAGGAACAAAAGCGAAACCAGGCTCAACTTACACTT 3125

Qy  2890 GGTTACTCACAAAGAACAACAAGAGTCAATGTACTTGTCCTAGCGTTTTGGAGAGAGAAAACA 2949
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3126 GGTTACTCACAAAGAACAACAAGAGTCAATGTACTTGTCCTAGCGTTTTGGAGAGAGAAAACA 3185
```

```
Qy  2950 GGAACCCACCAAACCAACCAATCAACCAAACAAGAAGAAAAATTCCACAATGAAAGAATGT 3009
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3186 GGAACCCACCAAACCAACCAATCAACCAAACAAGAAGAAAAATTCCACAATGAAAGAATGT 3245

Qy  3010 ATTTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTC 3069
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3246 ATTTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTC 3305

Qy  3070 TTT-AAAAAAAAAATGTGGAGAAAGTAGAATTTACCAAGGTTGTTGGCCCAGGGCGT 3128
         ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3306 TTTAAAAAAAAAATGTGGAGAAAGTAGAATTTACCAAGGTTGTTGGCCCAGGGCGT 3365

Qy  3129 TAAATTCACAGATTTTTTAACGAGAAAAACACACAGAAGAGCTACCTCAGTGTTTT 3188
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3366 TAAATTCACAGATTTTTTAACGAGAAAAACACACAGAAGAGCTACCTCAGTGTTTT 3425

Qy  3189 ACCTCAGCACCTTGCTCTTGTCTTGTCTTTCCCTTAGAGATTTGTAAAGCTGATAGTTGGAGCA 3248
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3426 ACCTCAGCACCTTGCTCTTGTCTTGTCTTTCCCTTAGAGATTTGTAAAGCTGATAGTTGGAGCA 3485

Qy  3249 TTTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAAAATAGAATATCAACTTGCCAGCCT 3308
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3486 TTTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAAAATAGAATATCAACTTGCCAGCCT 3545

Qy  3309 GGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTTCCGCTAGCCAAGA 3368
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3546 GGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTTCCGCTAGCCAAGA 3605

Qy  3369 ACC 3371
         |||
Db  3606 ACC 3608
```

RESULT 5
US-10-097-340-146
; Sequence 146, Application US/10097340
; Publication No. US20030087250A1
GENERAL INFORMATION:
 APPLICANT: John MONAHAN
 APPLICANT: Manjula GANNAVARAPU
 APPLICANT: Sebastian HOERSCH
 APPLICANT: Shubhangi KAMATKAR
 APPLICANT: Steve G. KOVATS
 APPLICANT: Rachel E. MEYERS
 APPLICANT: Michael MORRISEY
 APPLICANT: Peter OLANDT
 APPLICANT: Ami SEN
 APPLICANT: Peter VEIBY
 APPLICANT: Gordon B. MILLS
 APPLICANT: Robert C. BAST, Jr.
 APPLICANT: Karen LU
 APPLICANT: Rosemarie SCHMANDT
 APPLICANT: Xumei ZHAO
 APPLICANT: Karen GLATT
 TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
 TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer
 FILE REFERENCE: MRI-030
 CURRENT APPLICATION NUMBER: US/10/097,340
 CURRENT FILING DATE: 2002-03-14
 PRIOR APPLICATION NUMBER: 60/276,025
 PRIOR FILING DATE: 2001-03-14
 PRIOR APPLICATION NUMBER: 60/325,149
 PRIOR FILING DATE: 2001-09-26
 PRIOR APPLICATION NUMBER: 60/276,026
 PRIOR FILING DATE: 2001-03-14
 PRIOR APPLICATION NUMBER: 60/324,967
 PRIOR FILING DATE: 2001/09/26
 PRIOR APPLICATION NUMBER: 60/311,732
 PRIOR FILING DATE: 2001-08-10
 PRIOR APPLICATION NUMBER: 60/325,102
 PRIOR FILING DATE: 2001-09-26
 PRIOR APPLICATION NUMBER: 60/323,580
 PRIOR FILING DATE: 2001-09-19

```
;   NUMBER OF SEQ ID NOS: 363
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 146
;   LENGTH: 3667
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-097-340-146

Query Match           53.0%;   Score 1810;   DB 15;   Length 3667;
Best Local Similarity 99.9%;   Pred. No. 0;
Matches 2100;   Conservative 0;   Mismatches 2;   Indels 2;   Gaps 2;

Qy   1270 ACCCACTCCGATACTTCTCCGGTACCCCCATCACCAGTTGGCCCGTTCCCGCAT 1329
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   1507 ACCCACTCCGATACTTCTCCAGCCTGTACCCCCATCACCAGTTGGCCCGTTCCCGCAT 1566

Qy   1330 CATCACTCTTATCCAGAGCAGGAGAGATTGTGAATCTCTTCATCCCAACCCAGGCTGTGGGC 1389
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1567 CATCACTCTTATCCAGAGCAGGAGAGATTGTGAATCTCTTCATCCCAACCCAGGCTGTGGGC 1626

Qy   1390 GCCATCATCGGGAAGGAAGAGGGGCACACATCAAACAGCTGGCGAGATTCGCCGGAGCCTCT 1449
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1627 GCCATCATCGGGAAGGAAGAGGGGCACACATCAAACAGCTGGCGAGATTCGCCGGAGCCTCT 1686

Qy   1450 ATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCATCACCGGG 1509
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1687 ATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCATCACCGGG 1746

Qy   1510 CCACCGGAAGCCCAGTTCAAGGCCCAGGGACGGATCTCTTGGGAAACTGAAAGAGGAAAAC 1569
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1747 CCACCGGAAGCCCAGTTCAAGGCCCAGGGACGGATCTCTTGGGAAACTGAAAGAGGAAAAC 1806

Qy   1570 TTCTTTAACCCCAAAGAAGAAGAGTGAAAGTGGCAAGACCGTGAACGAACTGCAGAACTTAACCAGT 1629
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1807 TTCTTTAACCCCAAAGAAGAAGAGTGAAAGTGGCAAGACCGTGAACGAACTGCAGAACTTAACCAGT 1866

Qy   1630 GCTGGCCGGGTGATTGGCAAAGTGGCAAAGCCAGATGAAAATGAGGAAGTGATCGTCAGA 1689
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1867 GCTGGCCGGGTGATTGGCAAAGTGGCAAAGCCAGATGAAAATGAGGAAGTGATCGTCAGA 1926

Qy   1690 GCAGAAGTCATCGTCGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTGATCGTCAGA 1749
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1927 GCAGAAGTCATCGTCGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTGATCGTCAGA 1986

Qy   1750 ATTATCGGGCACTTCTTTTGCTAGCCAGACTGCACAGCCCAAGATCAGGGAAATTGTACAA 1809
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1987 ATTATCGGGCACTTCTTTTGCTAGCCAGACTGCACAGCCCAAGATCAGGGAAATTGTACAA 2046

Qy   1810 CAGGTGAAGCAGCAGGAGCGAGGAGAATAACCCTCAGGGAGTCGCCTCACAGCGCAGCAAGTGA 1869
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2047 CAGGTGAAGCAGCAGGAGCGAGGAGAATAACCCTCAGGGAGTCGCCTCACAGCGCAGCAAGTGA 2106

Qy   1870 GGCTCCCACCAGGCACCACCGGATGAAGTGAGCCTCCAAAGACCATCTGAGGAATGAGAA 1929
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2107 GGCTCCCACCAGGCACCACCGGATGAAGTGAGCCTCCAAAGACCATCTGAGGAATGAGAA 2166

Qy   1930 TGAGACCAAACGCAGCCAGCCAGTCGGGAGCAAACCAAAGACCATCTGAGGAATGAGAA 1989
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2167 TGAGACCAAACGCAGCCAGCCAGTCGGGAGCAAACCAAAGACCATCTGAGGAATGAGAA 2226

Qy   1990 GTCTGCGGAGGCGGCGCCCAGGGACTCTGCCGAGGGCCTGAGAACCCCAGGGGGCCGAGAGGG 2049
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   2227 GTCTGCGGAGGCGGCGCCCAGGGACTCTGCCGAGGGCCTGAGAACCCCAGGGGGCCGAGGAGGG 2286

Qy   2050 GCGGGGAAGGTCAGCCAGGTTTGCAACCACCGGGCCTCCGAGGGCCCCGCCCCAGGGC 2109
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2287 GCGGGGAAGGTCAGCCAGGTTTGCAACCACCGGGCCTCCGAGGGCCCCGCCCCAGGGC 2346

Qy   2110 TTCTGCAGGCTTCAGCCATCCACTTCACCATCGGATCTCGGATCTCCTGAACTCCCACGA 2169
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2347 TTCTGCAGGCTTCAGCCATCCACTTCACCATCGGATCTCGGATCTCCTGAACTCCCACGA 2406

Qy   2170 CGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGCAC 2229
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2407 CGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGCAC 2466

Qy   2230 ACCCTTTTCTGTTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAACGGGAAGATG 2289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2467 ACCCTTTTCTGTTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAAGGGGAAGATG 2526

Qy   2290 TTAAGATATGTGGCCTGGGTTACACAGGGGTGCCTGCAGCCGTAATATATTTAGAAAT 2349
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2527 TTAAGATATGTGGCCTGGGTTACACAGGGGTGCCTGCAGCCGTAATATATTTAGAAAT 2586

Qy   2350 AATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTTTTTTTCTTT 2409
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   2587 AATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTTTTTTTTCTTT 2646

Qy   2410 TTAAAGAGAAAGCAGCGTTTCTGAGACTTAAAGAATAAAGTCTTTGGGAGGTCTCACGG 2469
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2647 TTAAAGAGAAAGCAGCGTTTCTGAGACTTAAAGAATAAAGTCTTTGGGAGGTCTCACGG 2706

Qy   2470 TGTAGAGGAGCTTTGAGGCTGCACCCCCACAAAATTCACCCAGAGGAAATCTCGTCGGA 2529
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2707 TGTAGAGGAGCTTTGAGGCTGCACCCCCACAAAATTCACCCAGAGGAAATCTCGTCGGA 2766

Qy   2530 AGGACACTCACGGCAGTTCTGATCACCGTGTATGTCAACAGAGACGGGATACCGTCTCCT 2589
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2767 AGGACACTCACGGCAGTTCTGATCACCGTGTATGTCAACAGAGACGGGATACCGTCTCCT 2826

Qy   2590 TGAAGAGGAAACTCTGCACTCCTCATGCCTGCTCGAGCTCATCAGCTCATCACCCATTTCTCTTTTTGC 2649
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2827 TGAAGAGGAAACTCTGCACTCCTCATGCCTGCTCGAGCTCATCAGCTCATCACCCATTTCTCTTTTTGC 2886

Qy   2650 TTCACAGGTTTAAACTGGTTTTTGCATATCTGTATATAATTCTCGTCTCTCTCTGTT 2709
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2887 TTCACAGGTTTAAACTGGTTTTTGCATATCTGTATATAATTCTCGTCTCTCTCTGTT 2946

Qy   2710 TATCTCCCCTCCCTCCTTCTTCCATCTCCATTCTTTGAATTCCTCAT 2769
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   2947 TATCTCCCCTCCCTCCTTCTTCCATCTCCATTCTTTGAATTCCTCAT 3006

Qy   2770 CCCTCCATCTCAATCCCGTATCTACGGACCCCCCCCCCCAGGCAAAGCCAGCCTCAAACTTACCACTT 2829
          ||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   3007 CCCTCCATCTCAATCCCGTATCTACGGACCCCCCCCC-CCAGGCAAAGCCAGCCTCAAACTTACCACTT 3065

Qy   2830 GTATCACATCACACAAAAGGAACAAAAGCGAAACACACACAAACCACAAACACAACTTACACTT 2889
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3066 GTATCACATCACACAAAAGGAACAAAAGCGAAACACACACAAACCACAAACACAACTTACACTT 3125

Qy   2890 GGTTACTCAAAAGAACAAGAGTCAATGGTACTTGCCAGCGTTTGGAAGAGAGAAAACA 2949
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3126 GGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAGAGAGAAAACA 3185

Qy   2950 GGAACCCACCAAACCCAAACCAACCAATCAACAAAGAAAAATTCCACAATGAAAGAATGT 3009
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3186 GGAACCCACCAAACCCAAACCAACCAATCAACAAAGAAAAATTCCACAATGAAAGAATGT 3245

Qy   3010 ATTTTGTCTTTTGCATTTGGTGTATAAGCATCATTCAGCAGAAATGATTCCTTTC 3069
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3246 ATTTTGTCTTTTGCATTTGGTGTATAAGCATCATTCAGCAGAAATGATTCCTTTC 3305

Qy   3070 TTTAAAAAAAAATGTGGAGAAGTAGAAATTTACCAAGGTTGTTGCCCAGGGGCGT 3128
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3306 TTTAAAAAAAAATGTGGAGAAGTAGAAATTTACCAAGGTTGTTGCCCAGGGGCGCT 3365

Qy   3129 TAAATTCACAGATTTTTTAACGAAGAAAAACACAGAGAAGAAGCTACCCTCAGGTGTTTTT 3188
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3366 TAAATTCACAGATTTTTTTAACGAAGAAAAACACAGAGAAGAAGCTACCCTCAGGTGTTTTT 3425

Qy   3189 ACTTCAGCACCTTGCTTCTTGTTCTTGTTTCCCTTGAGATTTTTGTAAAGCTGATAGTTGGAGCA 3248
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3426 ACTTCAGCACCTTGATCTTGTGTTTTCCCTTGAGATTTTTGTAAAGCTGATAGTTGGAGCA 3485

Qy   3249 TTTTTTATTTTTAATAAAATGAGTTGAAAAAAATAAGATATCAACTGCCAGCCT 3308
```

Db   3486 TTTTTTATTTTTTTAATAAAAATGAGTTGGAAAAAAAAATAAGATATCAACTGCCAGCCT 3545
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3309 GGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTCTTCCGCTAGCCAAGA 3368

Db   3546 GGAGAAGGTGACAGTCCAAGTGTCAACAGCTGTCTTCCGCTAGCCAAGA 3605
           ||||||||
Qy   3369 ACC 3371

Db   3606 ACC 3608

RESULT 6
US-10-648-593-48
; Sequence 48, Application US/10648593
; Publication No. US2004010613A1
GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF GENES FOR PREDICTING ACTIVITY OF COMPOUNDS THAT
; TITLE OF INVENTION: INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE KINASES AND/OR
; TITLE OF INVENTION: PROTEIN TYROSINE KINASE PATHWAYS IN BREAST CELLS
; FILE REFERENCE: D0273 NP
; CURRENT APPLICATION NUMBER: US/10/648,593
; CURRENT FILING DATE: 2003-08-26
; PRIOR APPLICATION NUMBER: 60/406,385
; PRIOR FILING DATE: 2002-08-27
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 48
; LENGTH: 3667
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-648-593-48

Query Match   53.0%; Score 1810; DB 17; Length 3667;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2100; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy   1270 ACCCACTCCGATACTTCTCCAGCTGTACCCCATCACCAGTTTGGCCCGTTCCGCAT 1329
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1507 ACCCACTCCGATACTTCTCCAGCTGTACCCCATCACCAGTTTGGCCCGTTCCGCAT 1566

Qy   1330 CATCACTCTATCCAGAGCAGGAGGAGATTGTGAATCTCTTCAAGCATCCCCAACCCAGGCGTGTGGGC 1389
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1567 CATCACTCTATCCAGAGCAGGAGGAGATTGTGAATCTCTTCAAGCATCCCCAACCCAGGCGTGTGGGC 1626

Qy   1390 GCCATCATCGGGAAGAAGAAGGGGCACACATCAAAACGTCAGCGAATTCGCCGGAGCCTCT 1449
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1627 GCCATCATCGGGAAGAAGAAGGGGCACACATCAAAACAGCTCGGAGAATTCGCCGGAGCCTCT 1686

Qy   1450 ATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCATCACCGGG 1509
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1687 ATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCATCACCGGG 1746

Qy   1510 CCACCGGAAGCCCAGTTCAAGGCCCCAGGGACGATCTTTGGGAAACTGAAAGAGGAAAAC 1569
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1747 CCACCGGAAGCCCAGTTCAAGGCCCCAGGGACGATCTTTGGGAAACTGAAAGAGGAAAAC 1806

Qy   1570 TTCTTTAACCCCAAAGAAGAAGAGAGTGAAGCTGAAGCGCATATCAGAGTGCCCTCTTCCACA 1629
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1807 TTCTTTAACCCCAAAGAAGAAGAGAGTGAAGCTGAAGCGCATATCAGAGTGCCCTCTTCCACA 1866

Qy   1630 GCTGGCCCGGGTGATTGGCAAAGGTGGCAAGACCGTTGCAGAACTTAACCAGT 1689
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1867 GCTGGCCCGGGTGATTGGCAAAGGTGGCAAGACCGTTGCAGAACTTAACCAGT 1926

Qy   1690 GCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTTGATCGTCAGA 1749
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1927 GCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTTGATCGTCAGA 1986

Qy   1750 ATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCCGAAGTCAGCGGAAATTGTACAA 1809
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1987 ATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCCGAAGTCAGCGGAAATTGTACAA 2046

Db   1810 CAGGTGAAGCAGCAGCAGGAGCAGCAGAGGACAGCAGAAATACCCTCAGGGAGTCGCTCAGCGGAGCAGTGA 1869
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2047 CAGGTGAAGCAGCAGCAGGAGCAGCAGAGGACAGCAGAAATACCCTCAGGGAGTCGCTCAGCGGAGCAGTGA 2106

Qy   1870 GGCTCCCACAGCGCACCAGGGCACCGAGGACACCACCGAGCCCTTCCAACAACACTGACGAGAA 1929
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2107 GGCTCCCACAGCGCACCAGGGCACCGAGGACACCACCGAGCCCTTCCAACAACACTGACGAGAA 2166

Qy   1930 TGAGACCAAACGCCAGTCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAGGAATGAGAA 1989
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2167 TGAGACCAAACGCCAGTCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAGGAATGAGAA 2226

Qy   1990 GTCTGCGAGGCGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCAGGGGCCGAGGAGGG 2049
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2227 GTCTGCGAGGCGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCAGGGGCCGAGGAGGG 2286

Qy   2050 GCGGGGAAGGTCAGCCAGGTTTGCCAGGACACCAGCGAGCCCGCTCCCGGCCCCAGGGC 2109
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2287 GCGGGGAAGGTCAGCCAGGTTTGCCAGGACACCAGCGAGCCCGCTCCCGGCCCCAGGGC 2346

Qy   2110 TTCTGCAGGCTTCAGCCATCCACCATCCACCCGGATCTCTCCTGAACTCCCACGA 2169
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2347 TTCTGCAGGCTTCAGCCATCCACCATCCACCCGGATCTCTCCTGAACTCCCACGA 2406

Qy   2170 CGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTCAAAGCCAAGCAAGCAAAATGCAC 2229
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2407 CGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAGCAAAATGCAC 2466

Qy   2230 ACCCTTTTTCTGTGGCAAATGCTGCTCTGTACAGGGTGCTCGAGCGGTAATATATTTTAGAAATG 2289
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2467 ACCCTTTTTCTGTGGCAAATGCTGCTCTGTACAGGGTGCTCGAGCGGTAATATATTTTAGAAATG 2526

Qy   2290 TTAAGATATGTGGCCTGGTGGGTTACACAGGGTGCTGCCTGCAGCGGTAATATTAATTTTTTTTTTCTT 2349
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   2527 TTAAGATATGTGGCCTGGTGTGGGTTACACAGGGTGCTGCCTGCAGCGGTAATATTAATTTTTTTTTTCTT 2586

Qy   2350 AATATATCAAATAACTAACTAACTAACTCCAATTTTTAATCAATTATTAATTTTTTTTTTTCTT 2409
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2587 AATATATCAAATAACTAACTAACTAACTCCAATTTTTAATCAATTATTAATTTTTTTTTTTCTT 2646

Qy   2410 TTAAAGAAAGACAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTCACGG 2469
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2647 TTAAAGAAAGACAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGTGTCACGG 2706

Qy   2470 TGTAGAGGAGCTTTGAGGCCACCCGCACAAATTCACCCAGAGGGAAATCTCGTCGGA 2529
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2707 TGTAGAGGAGCTTTGAGGCCACCCGCACAAATTCACCCAGAGGGAAATCTCGTCGGA 2766

Qy   2530 AGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAAGGGATACCGTCTCCT 2589
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2767 AGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAAGGGATACCGTCTCCT 2826

Qy   2590 TGAAGAGGAAACTCTGTCACTCTGTCATGCCTGTCTAGCTCATACACCCATTTCTCTTTGC 2649
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2827 TGAAGAGGAAACTCTGTCACTCTGTCATGCCTGTCTAGCTCATACACCCATTTCTCTTTGC 2886

Qy   2650 TTCACAGGTTTAAACTGGTTTTTTGCATACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCTCTCTGTT 2709
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2887 TTCACAGGTTTAAACTGGTTTTTTGCATACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCTCTCTGTT 2946

Qy   2710 TATCTCTCCCCTCCCCTCCCCTCCCTCCTTCTTCTCTTCCATCTCCATTCTTTGAATTCCTCAT 2769
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2947 TATCTCTCCCCTCCCCTCCCCTCCCTCCTTCTTCTCTTCCATCTCCATTCTTTGAATTCCTCAT 3006

Qy   2770 CCCTCCATCTCAATCCCGTATCCATCCGTATCTACGCA CCCCCCCCCCCCAGGCAAAGCAGTGCTCTGA 2829
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3007 CCCTCCATCTCAATCCCGTATCCATCCGTATCTACGCA-CCCCCCCCCCCCAGGCAAAGCAGTGCTCTGA 3065

Qy   2830 GTATCACATCACACAAAGAGGCAAACACACAAAACCACAGCCTCCAACTTCCAACTT 2889
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3066 GTATCACATCACACAAAGAGGCAAACACACAAAACCACAGCCTCCAACTTCCAACTT 3125

Qy   2890 GGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTCAGCGTTTTGGAAGAGAAAACA 2949

```
Db          |||||||||||||||||||||||||||||||||||||||
       3126 GGTTACTCAAAGAACAAGAGTCAATGGTACTTGTGTCCTAGCGTTTTGGAAGAGGAAAACA 3185

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       2950 GGAACCCACCACCAACCAATCAACCAACCAAACAAAGAAAAAATTCCACATGAAAGAATGT 3009

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       3186 GGAACCCACCACCAACCAATCAACCAACCAAACAAAGAAAAAATTCCACATGAAAGAATGT 3245

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       3010 ATTTTGTCTTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTC 3069

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       3246 ATTTTGTCTTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTC 3305

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       3070 TTT-AAAAAAAAATGTGGAGGAAGTAGAAATTACCAAGGTTGTTGGCCCCAGGGGGT 3128

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       3306 TTTAAAAAAAAAATGTGGAGGAAGTAGAAATTACCAAGGTTGTTGGCCCCAGGGGCGT 3365

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       3129 TAAATTCACAGATTTTTTAACGAGAAAAAACACACAGAGAGAAGTCACCTCAGGTGTTTT 3188

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       3366 TAAATTCACAGATTTTTTAACGAGAAAAAACACACAGAGAGAAGTCACCTCAGGTGTTTT 3425

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       3189 ACCTCAGCACCTTGCTCTTGTGTTCCCTTAGAGATTTTGTAAAGCTCGTGATAGTTGGAGCA 3248

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       3426 ACCTCAGCACCTTGATCTTGTGTTCCCTTAGAGATTTTGTAAAGCTCGTGATAGTTGGAGCA 3485

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       3249 TTTTTTTATTTTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACTGCCAGCCT 3308

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       3486 TTTTTTTATTTTTTTTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACTGCCAGCCT 3545

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       3309 GGAGAAGGTGACAGTCCAAGTGTCCAAGCTGTTTCTTCCGGTAGCCAAGA 3368

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       3546 GGAGAAGGTGACAGTCCAAGTGTCCAAGCTGTGTCTTCCGGTTCTTCCGCTAGCCAAGA 3605

Qy          |||||
       3369 ACC 3371

Db          |||||
       3606 ACC 3608
```

RESULT 7
US-09-764-864-332
; Sequence 332, Application US/09764864
; Patent No. US2002013753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 332
; LENGTH: 2290
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-332

Query Match   43.8%; Score 1496; DB 9; Length 2290;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1966; Conservative 0; Mismatches 1; Indels 4; Gaps 3;

```
Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       1387 GGGCCATCATCGGGAAGAGGGGGCACACATCAAACAGCTGCGAGATTCGCCGGAGCC 1446

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
         36 GGGCCATCATCGGGAAGAGGGGGCACACATCAAACAGCTGCGAGATTCGCCGGAGCC 95

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       1447 TCTATCAAGATTGCCCCTGCGGAAGCGTCAGCGGGACGATGGTCATCATCACC 1506

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
         96 TCTATCAAGATTGCCCCTGCGGAAGCGTCAGCGGGACGATGGTCATCATCACC 155

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       1507 GGGCCACCGGAAGCCAGTTCAAGGCCCAGGCCCAGGAGACGGAGATCTTTGGGAAACTGAAAGAGGAA 1566

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
        156 GGGCCACCGGAAGCCAGTTCAAGGCCCAGGCCCAGGAGACGGAGATCTTTGGGAAACTGAAAGAGGAA 215

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       1567 AACTTCTTTAACCCCAAAGAAGAAGAGTGAAGCGCATATCAGAGTGCCCTCTTCC 1626

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       1296 TGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCTCT 1355
```

```
Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
        216 AACTTCTTTAACCCCAAAGAAGAAGAGTGAAGCGCATATCAGAGTGCCCTCTTCC 275

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       1627 ACAGCTGGCCGGGTGATTGGCAAGGTGGCAAGACCGTGAACGAACTGCAGAACTTAACC 1686

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
        276 ACAGCTGGCCGGGTGATTGGCAAGGTGGCAAGACCGTGAACGAACTGCAGAACTTAACC 335

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       1687 AGTGCAGAAGTCATCGTGCTGCCTCGTCGTGACCAAGACGCCAGAGAAATGAGGAAGTGATCGTC 1746

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
        336 AGTGCAGAAGTCATCGTGCTGCCTCGTCGTGACCAAACGCCAGATGAAATGAGGAAGTGATCGTC 395

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       1747 AGAATTATCGGGACCACTTCTTTGCTAGCCAGACTGCACAGCCGCAAGATCAGGGAAATTGTA 1806

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
        396 AGAATTATCGGGACCACTTCTTTGCTAGCCAGACTGCACAGCGGCAAGATCAGGGAAATTGTA 455

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       1807 CAACAGGTGAAGCAGCAGCAGGAGCAGAATACCCTCAGGGAGTGCGCCTCACAGCGCAGCAAG 1866

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
        456 CAACAGGTGAAGCAGCAGCAGGAGCAGAATACCCTCAGGGAGTGCGCCTCACAGCGCAGCAAG 515

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       1867 TGAGGCTCCCACAGGCACCAGCAGCCAGCAAACAACGATGAATGTAGCCCTTCCAACACCTGACA 1926

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
        516 TGAGGCTCCCACAGGCACCAGCAGCCAGCAAACAACGATGAATGTAGCCCTTCCAACACCTGACA 575

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       1927 GAATGAGACCAAACGCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAGGAATGA 1986

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
        576 GAATGAGACCAAACGCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAGGAATGA 635

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       1987 GAAGTCTGCGGAGGCGGCCAGGACTCTGCCAGGACTCTGCCGAGGACCACCGAGGGCCAGGA 2046

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
        636 GAAGTCTGCGGAGGCGGCCAGGACTCTGCCAGGACTCTGCCGAGGACCACCGAGGGCCAGGA 695

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       2047 GGGGCGGGAGGAAGGTCAGCCAGGTTTGCCAGAACCACCACCACTCCGGATCCTCCTGAACTCCCA 2106

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
        696 GGGGCGGGAGGAAGGTCAGCCAGGTTTGCCAGAACCACCACCACTCCGGATCCTCCTGAACTCCCA 755

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       2107 GGCTTCTGCAGGCTTCAGCCATCCACTTCACCATCCACTCGGATCTCTCCTGAACTCCCA 2166

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
        756 GGCTTCTGCAGGCTTCAGCCATCCACTTCACCATCCACTCGGATCTCTCCTGAACTCCCA 815

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       2167 CGACGCTATCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATG 2226

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
        816 CGACGCTATCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATG 875

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       2227 CACACCCTTTTTCTGTGTGGCAAATCGTCTCTGTACATGTGTGTACATGTGTACATTAGAAAGGGAAG 2286

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
        876 CACACCCTTTTTCTGTGTGGCAAATCGTCTCTGTACATGTGTGTACATGTGTACATTAGAAAGGGAAG 935

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       2287 ATGTTAAGATATGTGCCTGTGGGTTACACAGGGGTGCCTGCAGCCGGTAATAATATTTTAGA 2346

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
        936 ATGTTAAGATATGTGCCTGTGGGTTACACAGGGGTGCCTGCAGCCGGTAATAATATTTTAGA 995

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       2347 AATAATAATATCAAATAACTCAACTAACTCAATTTTTAATCAATTATTTAATTAATTTTTTTTTTC 2406

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
        996 AATAATAATATCAAATAACTCAACTAACTCAATTTTTAATCAATTATTTAATTAATTTTTTTTTTTC 1055

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       2407 TTTTTAAAGAGAAAGCAGGCTTTTGAGGCTTTTCTAGACTTAAAGACTAAAGTCTTTGGGAGGTCTCA 2466

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       1056 TTTTTAAAGAGAAAGCAGGCTTTTGAGGCTTTTCTAGACTTAAAGACTAAAGTCTTTGGGAGGTCTCA 1115

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       2467 CGGTGTAGAGAGAGGAGCTTTGAGGCCACCCCACAAAATTCACCCGAGGGAAATCTCGTC 2526

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       1116 CGGTGTAGAGAGAGGAGCTTTGAGGCCACCCCACAAAATTCACCCGAGGGAAATCTCGTC 1175

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       2527 GGAAGGACACTCACCGCAGTTCTGACAGCAGTTCTGGAACAGAAGGATCCGGATACCGTCT 2586

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       1176 GGAAGGACACTCACCGCAGTTCTGACAGCAGTTCTGGAACAGAAGGATACCGTCT 1235

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       2587 CCTTGAAGAGGAAACTTCTGTCACTTCTGTCTATGCCTTGTCTAGCTGTCTCATACACCCATTTCTCTT 2646

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       1236 CCTTGAAGAGGAAACTTCTGTCACTCGTCTATGCCTTGTCTAGCTGTCTCATACACCCATTTCTCTT 1295

Qy          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       2647 TGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCTCT 2706

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||
       1296 TGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCTCT 1355
```

US-09-764-864-749

```
Query Match             11.8%;  Score 401;  DB 9;  Length 1186;
Best Local Similarity   99.4%;  Pred. No. 3.2e-194;
Matches  671;  Conservative   0;  Mismatches   3;  Indels   1;  Gaps   1;

Qy   1864 AAGTGAGGCTCCACAGCACCAGCAAACACGGATGAATGTAGCCTTCCAACACCTG 1923
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    512 AAGTGAGGCTCCACAGCACCAGCAAACACGGATGAATGTAGCCCTTCCAACACCTG 571

Qy   1924 ACAGAATAGAGACCAAAGCCAGCCAGCCCAGATGGGAGCAAACCAAAGACCATCTGAGGAA 1983
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    572 ACAGAATAGAGACCAAAGCCAGCCAGCCCAGATGGGAGCAAACCAAAGACCATCTGAGGAA 631

Qy   1984 TGAGAAGTCTGCGGAGGCGGCGGCCCAGGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGCCGA 2043
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    632 TGAGAAGTCTGCGGAGGCGGCGGCCCAGGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGCCGA 691

Qy   2044 GGAGGGGCGGGGAAGGTCAGCCAGGTTGCCAGAACCACCGAGCCCCGGCTCCGCGCCCCC 2103
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    692 GGAGGGGCGGGGAAGGTCAGCCAGGTTGCCAGAACCACCGAGCCCCGGCTCCGCGCCCCC 751

Qy   2104 CAGGGCTTCGCAGGCTTCAGCCAT-CCACTTCACCATCCACTTCCGGATCTCTCCTGAACT 2162
          ||||||||||||||||||| | ||||||||||||||||||||||||||||||
Db    752 CAGGGCTTCGCAGGCTTCAGCNATNCCACTTCACCATCCACTTCCGGATCTCTCCTGAACT 811

Qy   2163 CCCACGACGCTATCCCCTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGCAA 2222
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    812 CCCACGACGCTATCCCTTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGCAA 871

Qy   2223 AATGCACACCCCTTTTCTGTGGCAAATCGTCTCTGTACATAGTGTACATATTAGAAAGG 2282
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    872 AATGCACACCCCTTTTCTGTGGCAAATCGTCTCTGTACATAGTGTACATATTAGAAAGG 931

Qy   2283 GAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGTGCCTGCAGCGGGTAATATATTT 2342
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    932 GAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGTGCCTGCAGCGGGTAATATATTT 991

Qy   2343 TAGAAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATATTAATAATTTTTT 2402
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    992 TAGAAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATATTAATCAATTTTTT 1051

Qy   2403 TTTCTTTTTAAAGAGAAAGCAGCGTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGT 2462
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1052 TTTCTTTTTAAAGAGRAAGCMGGCTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGT 1111

Qy   2463 CTCACGGTGTAGAGAGGACACTCCGCACCCGCACCCGCACAAAATTCACCCAGAGGGAAATCT 2522
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1112 CTCACGGTGTAGAGAGGACTTTGAGGCGCACCCGCACCCGCACAAAATTCACCCAGAGGGAAATCT 1171

Qy   2523 CGTCGGAAGGACACT 2537
          |||||||||||||
Db   1172 CGTCGGAAGGACACT 1186
```

RESULT 8
US-09-764-864-749
; Sequence 749, Application US/09764864
; Patent No. US2002013732753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 749
; LENGTH: 1186
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (511)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (774)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (777)
; OTHER INFORMATION: n equals a,t,g, or c

```
Qy   2707 GTTTATCTCTCCCCTCCCTCCCCCTTCTTCTCCATCTCCATTCTTTGAATTCCT 2766
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1356 G-TTATCTCTCCCCTCCCTCCCCCTTCTTCTCCATCTCCATTCTTTGAATTCCT 1414

Qy   2767 CATCCCTCCATCTCAATCCGTATCTACGGA--CCCCCCCCCCCCCCCAGGCAAAGCAGTGC 2824
          ||||||||||||||||||||||||||  ||||||||||||||||||||
Db   1415 CATCCCTCCATCTCAATCCGTATCTACGCACCCCCCCCCCCCCCCAGGCAAAGCAGTGC 1474

Qy   2825 TCTGAGTATCACATCACACAAAAGGAACAAAAGGCGAAACACACAAACCAGCCTCAACTTA 2884
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1475 TCTGAGTATCACATCACACAAAAGGAACAAAAGGCGAAACACACAAACCAGCCTCAACTTA 1534

Qy   2885 CACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAGAGGA 2944
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1535 CACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAGAGGA 1594

Qy   2945 AAACAGGAACCCACCAAACCAAACCAATCAACCAAACAAGAAAAAATTCCACAATGAAAG 3004
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1595 AAACAGGAACCCACCAAACCAAACCAATCAACCAAACAAGAAAAAATTCCACAATGAAAG 1654

Qy   3005 AATGTATTTTGTCTTTTTGCATTTTGGTGTATAAGCCATCAATTTCAGCAAAATGATTC 3064
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1655 AATGTATTTTGTCTTTTTGCATTTTGGTGTATAAGCCATCAATTTCAGCAAAATGATTC 1714

Qy   3065 CTTTCTTT-AAAAAAAAAAATGTGGAGGAAGTAGAAATTTACCAAGGTTGTTGGCCCAG 3123
          ||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1715 CTTTCTTTAAAAAAAAAAAATGTGGAGGAAGTAGAAATTTACCAAGGTTGTTGGCCCAG 1774

Qy   3124 GGCGTTAAATTCACAGATTTTTTAAGGAGAAAAACACACAGAAGAAGCTACCTCAGGTG 3183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1775 GGCGTTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAAGAAGCTACCTCAGGTG 1834

Qy   3184 TTTTTACCTCAGCACCTTGCTCTTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTTG 3243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1835 TTTTTACCTCAGCACCTTGCTCTTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTTG 1894

Qy   3244 GAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACTGCC 3303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1895 GAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACTGCC 1954

Qy   3304 AGCCTGGAGAGGTGACAGTGCCAGTGTGCAACAGCTGTTCTGAATTGTCT 3354
          ||||||||||||||||||||||||||||||||||||||||||||
Db   1955 AGCCTGGAGAGGTGACAGTGCCAGTGTGCAACAGCTGTTCTGAATTGTCT 2005
```

RESULT 9
US-09-814-353-14725
; Sequence 14725, Application US/09814353
; Publication No. US2003016581A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25

```
;   PRIOR APPLICATION NUMBER: US 60/211,940
;   PRIOR FILING DATE: 2000-06-15
;   PRIOR APPLICATION NUMBER: US 60/216,820
;   PRIOR FILING DATE: 2000-07-07
;   PRIOR APPLICATION NUMBER: US 60/220,661
;   PRIOR FILING DATE: 2000-07-25
;   PRIOR APPLICATION NUMBER: US 60/257,672
;   PRIOR FILING DATE: 2000-12-21
;   NUMBER OF SEQ ID NOS: 22037
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 14725
;   LENGTH: 582
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-814-353-14725

Query Match             11.6%;   Score 397;  DB 10;  Length 582;
Best Local Similarity   99.8%;   Pred. No. 3.5e-192;
Matches   517;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy   2855  AAGCGAAACACACAAACCAGCGCTCAACTTACACTTGGTTACTCAAAAGAACAAGAGTCAA  2914
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21  AAGCGAAACACACAAACCAGCGCTCAACTTACACTTGGTTACTCAAAAGAACAAGAGTCAA    80

Qy   2915  TGGTACTTGTCCTAGCGTTTTGGAAGAGAGAAAACAGGAAACCCACCAACCAACCAATCAA  2974
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81  TGGTACTTGTCCTAGCGTTTTGGAAGAGAGAAAACAGGAAACCCACCAACCAACCAATCAA   140

Qy   2975  CCAAACAAAGAAAAAAATTCCACAATGAAAGAATGTATTTTGTCTTTTTGCATTTTGGTGT  3034
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141  CCAAACAAAGAAAAAAATTCCACAATGAAAGAATGTATTTTGTCTTTTTGCATTTTGGTGT   200

Qy   3035  ATAAGCCATCAATATTCAGCAAAATGATTCCTTTCTT-AAAAAAAAAATGTGGAGAA  3093
                 ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    201  ATAAGCCATCAATATTCAGCAAAATGATTCCTTTCTTCTTAAAAAAAAAATGTGGAGAA   260

Qy   3094  AGTAGAAATTTACCAAGGTTGTTGGCCCAGGGCGTTAAATTCACAGATTTTTTTAACGAG  3153
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261  AGTAGAAATTTACCAAGGTTGTTGGCCCAGGGCGTTAAATTCACAGATTTTTTTAACGAG   320

Qy   3154  AAAAACACACAGAGAGAAGCTACCTCAGGTGTTTTACCTCAGCACCTTGCTCTTGTGTTT  3213
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321  AAAAACACACAGAGAGAAGCTACCTCAGGTGTTTTACCTCAGCACCTTGCTCTTGTGTTT   380

Qy   3214  CCCTTAGAGATTTTGTAAAGCTGAATAGTTGGAGCATTTTTTATTTTTTTAATAAAAATG  3273
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381  CCCTTAGAGATTTTGTAAAGCTGAATAGTTGGAGCATTTTTTATTTTTTTAATAAAAATG   440

Qy   3274  AGTTGGAAAAAAAAAATAAGATATCAACTGCCAGCCTGGAGAAGAAGGTGACAGTCCAAGTGTGC  3333
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    441  AGTTGGAAAAAAAAAATAAGATATCAACTGCCAGCCTGGAGAAGAAGGTGACAGTCCAAGTGTGC   500

Qy   3334  AACAGCTGTTCTGAATTGTCTTCCGCTAGCCAAGAACC  3371
                 ||||||||||||||||||||||||||||||||||||||
Db    501  AACAGCTGTTCTGAATTGTCTTCCGCTAGCCAAGAACC   538

RESULT 10
US-09-764-330
;  Sequence 330, Application US/09764864
;  Patent No. US2002013275A1
;  GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PT223
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE: 2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
;  SEQ ID NO 330
;  LENGTH: 822
;  TYPE: DNA
```

```
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (441)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (455)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-330

Query Match             11.4%;   Score 389;  DB 9;  Length 822;
Best Local Similarity   99.8%;   Pred. No. 4.5e-188;
Matches   439;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    779  CTGCAGAAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCA   838
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  CTGCAGAAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCA    60

Qy    839  TGATTCTTGAAATCATCGAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTC   898
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  TGAAATTCTTGAAATCATCGAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTC   120

Qy    899  TGAAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAGAGGCAGAAATT   958
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  TGAAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAGAGGCAGAAATT   180

Qy    959  TGAAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGA  1018
                 ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    181  TGAAGAAAATTGAACATGAAACAGGGACCAAGRTAACAATCTCATCTTTGCAGGATTTGA   240

Qy   1019  GCATATACAACCCGAAAGAACCATCACTGTGAAGGCCACAGTTGAGGCCTGTGCCAGTG  1078
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  GCATATACAACCCGAAAGAACCATCACTGTGAAGGCCACAGTTGAGGCCTGTGCCAGTG  1078

Qy   1079  CTGAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATGATGATGTGTTA  1138
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  CTGAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATGATGATGTGTTA   360

Qy   1139  ACCAACAAGCCAATCGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTCAACAG  1198
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  ACCAACAAGCCAATCGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTCAACAG   420

Qy   1199  GACTGTCCGTGCTATCTCCA  1218
                 ||||||||||||||||||||
Db    421  GACTGTCCGTGCTATCTCCA   440

RESULT 11
US-09-814-353-1995
;  Sequence 1995, Application US/0981453
;  Publication No. US20030165831A1
;  GENERAL INFORMATION:
;  APPLICANT: Lee, John
;  APPLICANT: Thompson, Pamela
;  APPLICANT: Lillie, James
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
;  FILE REFERENCE: MRI-006B
;  CURRENT APPLICATION NUMBER: US/09/814,353
;  CURRENT FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/191,031
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: US 60/207,124
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: US 60/211,940
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: US 60/216,820
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: US 60/220,661
;  PRIOR FILING DATE: 2000-07-25
;  PRIOR APPLICATION NUMBER: US 60/257,672
;  PRIOR FILING DATE: 2000-12-21
```

Left column

```
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1995
; LENGTH: 579
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-1995

Query Match        10.2%;      Score 347;   DB 10;   Length 579;
Best Local Similarity   99.8%;      Pred. No. 1.5e-166;
Matches 467;   Conservative   0;   Mismatches   0;   Indels   1;   Gaps   1;

Qy 2846 AAGGAACAAAAGCGAAACACACAAACCAGCCTCAACTTACACTTGGTTACTCAAAAGAAC 2905
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8 AAGGAACAAAAGCGAAACACACAAACCAGCCTCAACTTACACTTGGTTACTCAAAAGAAC 67

Qy 2906 AAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAGAGGAAAACAGGAACCCACCAAACCA 2965
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68 AAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAGAGGAAAACAGGAACCCACCAAACCA 127

Qy 2966 ACCAATCAACCAAACAAAGAAAAAATTCCACAATGAAAGAATGTATTTTGTCTTTTTGCA 3025
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  128 ACCAATCAACCAAACAAAGAAAAAATTCCACAATGAAAGAATGTATTTTGTCTTTTTGCA 187

Qy 3026 TTTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTCTT-AAAAAAAAAT 3084
        ||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||
Db  188 TTTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTCTTTAAAAAAAAAAT 247

Qy 3085 GTGGAGGAAAGTAGAAATTACCAAGGTGTTGGCCCAGGGCGTTAAATTCACAGATTT 3144
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  248 GTGGAGGAAAGTAGAAATTACCAAGGTGTTGGCCCAGGGCGTTAAATTCACAGATTT 307

Qy 3145 TTTAACGAGAAAAACACACAGAAGAAGCTACCTCAGGTGTTTTTACCTCAGCACCTTGCT 3204
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  308 TTTAACGAGAAAAACACACAGAAGAAGCTACCTCAGGTGTTTTTACCTCAGCACCTTGCT 367

Qy 3205 CTTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTAGTTGGAGCATTTTTTATTTTTTA 3264
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  368 CTTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTAGTTGGAGCATTTTTTATTTTTTA 427

Qy 3265 ATAAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCAGCCTGGAG 3312
        |||||||||||||||||||||||||||||||||||||||||||||||
Db  428 ATAAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCAGCCTGGAG 475
```

RESULT 12
US-09-814-353-8341
; Sequence 8341, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25

Right column

```
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8341
; LENGTH: 579
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-8341

Query Match        10.2%;      Score 347;   DB 10;   Length 579;
Best Local Similarity   99.8%;      Pred. No. 1.5e-166;
Matches 467;   Conservative   0;   Mismatches   0;   Indels   1;   Gaps   1;

Qy 2846 AAGGAACAAAAGCGAAACACACAAACCAGCCTCAACTTACACTTGGTTACTCAAAAGAAC 2905
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8 AAGGAACAAAAGCGAAACACACAAACCAGCCTCAACTTACACTTGGTTACTCAAAAGAAC 67

Qy 2906 AAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAGAGGAAAACAGGAACCCACCAAACCA 2965
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68 AAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAGAGGAAAACAGGAACCCACCAAACCA 127

Qy 2966 ACCAATCAACCAAACAAAGAAAAAATTCCACAATGAAAGAATGTATTTTGTCTTTTTGCA 3025
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  128 ACCAATCAACCAAACAAAGAAAAAATTCCACAATGAAAGAATGTATTTTGTCTTTTTGCA 187

Qy 3026 TTTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTCTT-AAAAAAAAAT 3084
        ||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||
Db  188 TTTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTCTTTAAAAAAAAAAT 247

Qy 3085 GTGGAGGAAAGTAGAAATTACCAAGGTGTTGGCCCAGGGCGTTAAATTCACAGATTT 3144
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  248 GTGGAGGAAAGTAGAAATTACCAAGGTGTTGGCCCAGGGCGTTAAATTCACAGATTT 307

Qy 3145 TTTAACGAGAAAAACACACAGAAGAAGCTACCTCAGGTGTTTTTACCTCAGCACCTTGCT 3204
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  308 TTTAACGAGAAAAACACACAGAAGAAGCTACCTCAGGTGTTTTTACCTCAGCACCTTGCT 367

Qy 3205 CTTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTAGTTGGAGCATTTTTTATTTTTTA 3264
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  368 CTTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTAGTTGGAGCATTTTTTATTTTTTA 427

Qy 3265 ATAAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCAGCCTGGAG 3312
        |||||||||||||||||||||||||||||||||||||||||||||||
Db  428 ATAAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCAGCCTGGAG 475
```

RESULT 13
US-10-066-543-2949
; Sequence 2949, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2949
; LENGTH: 472
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-066-543-2949

Query Match        9.8%;  Score 333;  DB 15;  Length 472;
Best Local Similarity  99.8%;  Pred. No. 2.2e-159;
Matches  453;  Conservative  0;  Mismatches  0;  Indels  1;  Gaps  1;

Qy   2919 ACTTGTCCTAGCGTTTGGAAGAGAGAAAACAGGAACCCACCAAACCAACCAATCAACCAA 2978
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ACTTGTCCTAGCGTTTGGAAGAGAGAAAACAGGAACCCACCAAACCAACCAATCAACCAA 60

Qy   2979 ACAAAGAAAAAATTCCACAATGAAAGAATGTATTTTGTCTTTTTGCATTTTGGTGTATAA 3038
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACAAAGAAAAAATTCCACAATGAAAGAATGTATTTTGTCTTTTTGCATTTTGGTGTATAA 120

Qy   3039 GCCATCAATATTCAGCAAAATGATTCCTTTCTTT-AAAAAAAAATGTGGAGGAAAGTA 3097
          ||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db    121 GCCATCAATATTCAGCAAAATGATTCCTTTCTTTCTTTAAAAAAAAATGTGGAGGAAAGTA 180

Qy   3098 GAAATTTACCAAGGTGTTGGCCCAGGGCGTTAAATTCACAGATTTTTTAACGAGAAAA 3157
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GAAATTTACCAAGGTGTTGGCCCAGGGCGTTAAATTCACAGATTTTTTAACGAGAAAA 240

Qy   3158 ACACACAGAAGAAGCTACCTCAGTGTTTTCACCTCAGCACCTTGCTCTTGTTTCCCT 3217
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ACACACAGAAGAAGCTACCTCAGTGTTTTCACCTCAGCACCTTGCTCTTGTTTCCCT 300

Qy   3218 TAGAGATTTTGTAAAGCTGATAGTTGGAGCATTTTTTATTTTTTAATAAAAATGAGTT 3277
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TAGAGATTTTGTAAAGCTGATAGTTGGAGCATTTTTTATTTTTTAATAAAAATGAGTT 360

Qy   3278 GGAAAAAAAATAAGATATCAACTGCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACA 3337
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGAAAAAAAATAAGATATCAACTGCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACA 420

Qy   3338 GCTCGTTCTCGAATTGTCTTCCGCTAGCCAAGAACC 3371
          ||||||||||||||||||||||||||||||||||||
Db    421 GCTCGTTCTCGAATTGTCTTCCGCTAGCCAAGAACC 454

RESULT 14
US-09-777-564-1124
; Sequence 1124, Application US/09777564
; Patent No. US2002022591A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.493
; CURRENT APPLICATION NUMBER: US/09/777,564
; CURRENT FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 1730
; SOFTWARE: FastSEQ for Window Version 4.0
; SEQ ID NO 1124
; LENGTH: 264
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-777-564-1124

Query Match        7.7%;  Score 264;  DB 9;  Length 264;
Best Local Similarity  100.0%;  Pred. No. 4.9e-124;
Matches  264;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   3075 AAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTGTTGTTGGCCCAGGGCGTTAAATT 3134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTGTTGTTGGCCCAGGGCGTTAAATT 60

Qy   3135 CACAGATTTTTTAACGAGAAAACACACAGAAGAAGCTACCTCAGTGTTTTTACCTCA 3194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     61 CACAGATTTTTTAACGAGAAAACACACAGAAGAAGCTACCTCAGTGTTTTTACCTCA 120

Qy   3195 GCACCTTGCTCTTGTGTTTCCCTTAGAGATTTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGGAGCATTTTT 3254
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCACCTTGCTCTTGTGTTTCCCTTAGAGATTTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGGAGCATTTTT 180

Qy   3255 TATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCAGCCTGGAGAA 3314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCAGCCTGGAGAA 240

Qy   3315 GGTGACAGTCCAAGTGTGCAACAG 3338
          ||||||||||||||||||||||||
Db    241 GGTGACAGTCCAAGTGTGCAACAG 264

RESULT 15
US-10-015-219-1124
; Sequence 1124, Application US/10015219
; Publication No. US2003000829A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.493C1
; CURRENT APPLICATION NUMBER: US/10/015,219
; CURRENT FILING DATE: 2002-03-02
; NUMBER OF SEQ ID NOS: 1739
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1124
; LENGTH: 264
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-015-219-1124

Query Match        7.7%;  Score 264;  DB 15;  Length 264;
Best Local Similarity  100.0%;  Pred. No. 4.9e-124;
Matches  264;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   3075 AAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTGTTGTTGGCCCAGGGCGTTAAATT 3134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTGTTGTTGGCCCAGGGCGTTAAATT 60

Qy   3135 CACAGATTTTTTAACGAGAAAACACACAGAAGAAGCTACCTCAGTGTTTTTACCTCA 3194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CACAGATTTTTTAACGAGAAAACACACAGAAGAAGCTACCTCAGTGTTTTTACCTCA 120

Qy   3195 GCACCTTGCTCTTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGGAGCATTTTT 3254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCACCTTGCTCTTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGGAGCATTTTT 180

Qy   3255 TATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCAGCCTGGAGAA 3314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCAGCCTGGAGAA 240

Qy   3315 GGTGACAGTCCAAGTGTGCAACAG 3338
          ||||||||||||||||||||||||
Db    241 GGTGACAGTCCAAGTGTGCAACAG 264

Search completed: July 21, 2004, 23:22:09
Job time : 1438 secs
```

```
                    GenCore version 5.1.6
              Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:        July 16, 2004, 10:30:59 ; Search time 53.5 Seconds
                                  (without alignments)
                              12269.364 Million cell updates/sec

Title:          US-09-270-437D-6
Perfect score:  6121
Sequence:       1 ggcacgcaggaggccgagga.........aaccttgaaatgttattt 3412

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters:     566732

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries
```

```
Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09270437/runat_16072004_113050_13341/app_query.fasta_1.3591
-DB=PIR_78 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=Bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0927O437 @CGN_1_1_86 @runat_16072004_113050_13341 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=IC0 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  PIR_78:*
     1:  pir1:*
     2:  pir2:*
     3:  pir3:*
     4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	494	8.1	854	2	T23837	hypothetical prote
2	279.5	4.6	568	2	T49962	hypothetical prote
3	277.5	4.5	398	2	T41600	probable pre-mRNA
4	269	4.4	621	2	D96554	hypothetical prote
5	265.5	4.3	680	2	T25832	hypothetical prote
6	259	4.2	768	2	T27855	hypothetical prote
7	254.5	4.2	589	2	T19216	hypothetical prote
8	254.5	4.2	611	2	T19217	hypothetical prote
9	244	4.0	644	2	A53184	myc far upstream e
10	238.5	3.9	356	2	S58529	alpha-complex prot
11	236.5	3.9	846	2	T04533	hypothetical prote
12	235	3.8	413	2	S46109	hnRNP complex prot
13	226	3.7	479	2	C86275	hypothetical prote
14	226	3.7	510	2	I38489	onconeural ventral
15	225.5	3.7	365	2	S42471	hnRNP protein E2 -
16	225	3.7	362	2	S78515	single-stranded nu
17	223.5	3.6	632	2	T02627	hypothetical prote
18	219.5	3.6	313	2	T48439	probable RNA-bindi
19	217.5	3.5	397	2	T30168	hypothetical prote
20	209.5	3.4	396	2	S41224	hnRNP protein - Af
21	203.5	3.3	1268	2	A44125	high density lipop
22	201.5	3.3	1270	2	S23464	vigilin - chicken
23	198	3.2	1279	2	T41389	rna binding protei
24	194.5	3.2	748	2	T04011	hypothetical prote
25	193.5	3.2	1220	2	T19117	hypothetical prote
26	192.5	3.1	1198	2	B88279	protein C08H9.2 [i
27	186.5	3.0	464	2	S43363	transformation upr
28	186.5	3.0	464	2	A54143	kappa-B motif-bind
29	185.5	3.0	463	2	S41495	dC stretch-binding
30	185.5	3.0	1684	2	JW0057	gravin - human
31	184.5	3.0	629	2	T05425	polyadenylate-bind
32	184.5	3.0	806	2	T13690	hypothetical prote
33	182	2.9	968	2	T00353	hypothetical prote
34	179	2.9	924	2	S27923	gene LF3 protein -
35	179	2.9	1283	2	T49692	related to SCP160
36	179	2.9	4957	2	T03455	ALR protein - huma
37	179	2.9	5262	2	T03454	ALR protein - huma
38	178.5	2.9	473	2	A54691	octamer-binding pr
39	177.5	2.9	471	2	S41768	splicing factor ho
40	176.5	2.9	471	2	G01211	54 kDa protein - h
41	176	2.9	1222	2	S56030	SCP160 protein - y
42	175.5	2.9	1791	2	T02345	hypothetical prote
43	174.5	2.9	817	2	S51342	verprolin - yeast
44	168.5	2.8	2187	2	T30826	nascent polypeptid
45	167	2.7	638	2	S37085	polyadenylate-bind

ALIGNMENTS

```
RESULT 1
T23837
hypothetical protein M88.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T23837
R;Sulston,J.
submitted to the EMBL Data Library, June 1994
A;Reference number: Z19806
A;Accession: T23837
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-854 <WIL>
A;Cross-references: EMBL:Z34802; PIDN:CAA84338.1; GSPDB:GN00021; CESP:M88.5
A;Experimental source: clone M88
C;Genetics:
C;Gene: CESP:M88.5
A;Map position: 3
A;Introns: 38/3; 96/1; 133/3; 255/3; 307/3; 395/2; 436/2; 492/3; 657/3; 759/1; 820/1

Alignment Scores:
Pred. No.:
Score:                   494.00          Length:         854
Percent Similarity:      41.86%          Matches:        161
Best Local Similarity:   26.22%          Conservative:   96
Query Match:             8.07%           Mismatches:     182
DB:                                      Indels:         176
                                         Gaps:           21

US-09-270-437D-6 (1-3412) x T23837 (1-854)

Qy  422 CAGAAACCGCCGTTGTCAACGTCACCATATGCAACAAGA----------------------- 459
                 |||  |:||
Db  244 GlnHisProGlnMetMetGlnGlnHisAlaGlnGlnGlyTyrHisProHisGlnGlnAsn 263

Qy  460 GAAGAAGCAAAAATAGCCATGGAGAAGAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTC 519
             :::  |||||||  :::  ||||||  |||||  |||| ||||
Db  264 GlnGlnHisGlnAlaGlyGlnHisGlnSerHisGlnSerGlnAsnHisAsn--- 282
```

The page consists of a nucleotide/protein sequence alignment between a Query (Qy) and Database (Db) sequence, shown in paired rows with numeric position markers, followed by a RESULT block.

```
RESULT 2
T49962
hypothetical protein F8M21.160 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C;Accession: T49962
R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Le...
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24493
A;Accession: T49962
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-568 <BEV>
A;Cross-references: EMBL:AL353993; GSPDB:GN00063; ATSP:F8M21.160
A;Experimental source: cultivar Columbia; BAC clone F8M21
C;Genetics:
C;Gene: ATSP:F8M21.160
A;Map position: 5
A;Introns: 200/3; 337/3; 544/3

Alignment Scores:
Pred. No.:                                            Length:            568
Score:                 3.52e-11                        Matches:           140
                       279.50                          Conservative:      100
Percent Similarity:    40.82%                          Mismatches:        242
Best Local Similarity: 23.81%                          Indels:            107
Query Match:           4.57%                           Gaps:              24
DB:                    2

US-09-270-437D-6 (1-3412) x T49962 (1-568)
```

```
RESULT 3
T41600
probable pre-mRNA splicing factor - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T41600
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Bothe, G.; Pohl, T.
submitted to the EMBL Data Library, October 1998
A;Reference number: Z22003
A;Accession: T41600
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-398 <LYN>
A;Cross-references: EMBL:AL031825; PIDN:CAA21234.1; GSPDB:GN00068; SPDB:SPCC757.09c
A;Experimental source: strain 972h-; cosmid c757
C;Genetics:
A;Gene: SPDB:SPCC757.09c
A;Map position: 3

Alignment Scores:
Pred. No.:                              Length:         398
Score:                   4.38e-11       Matches:        108
Percent Similarity:      41.38%         Conservative:   60
Best Local Similarity:   26.60%         Mismatches:     157
Query Match:             4.53%          Indels:         81
DB:                      2              Gaps:           15
```

```
Db    327  IleProAlaAspMetValGlyCysIleIleGlyArgGlyGlySerLysIleSerGluIle  346

Qy   1429  GCCAGATTCGCCGAGCCTCTATCAAGATTGCCCTCGGAAGGCCCAGACCTCAGCGAA  1488
           :::||||:::  |||  ||||||  |||||:|||
Db    347  ArgArgThrSerGlySerLysIleSerIleAlaLysGluProHisAspGluThrGlyGlu  366

Qy   1489  AGGATGGTCATCATCACGGGCCACGGAAGCCCAGTTCAAGGCCCAGGACGGATCTTT  1548
           |||||
Db    367  ArgMetPheThrIleThrGlyThrHisGluGluAsnGluLysAlaLeuPheLeuLeuTyr  386

Qy   1549  GGGAAACTGAAAGAGGAA  1566
           :::||||:::
Db    387  GlnGlnLeuGluMetGlu  392
```

RESULT 4
D96554
hypothetical protein F19C24.19 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: D96554
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
J.; Chung, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: D96554
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-621 <STO>
A;Cross-references: GB:AE005173; NID:g11094762; NID:g11094762; PIDN:AAG29695.1; GSPDB:GN00141
C;Genetics:
A;Gene: F19C24.19
A;Map position: 1

Alignment Scores:
Pred. No.:   621
Score: 1.9e-10   Length: 621
                                        Matches: 134
Percent Similarity: 37.37%   Conservative: 88
Best Local Similarity: 22.56%   Mismatches: 206
Query Match: 4.39%   Indels: 166
DB: 2   Gaps: 26

US-09-270-437D-6 (1-3412) x D96554 (1-621)
```

```
Qy    616  CCTGGGGGCACTTCTCAGGCCAGACAGATTGATTCCCGTGCGGATCGTGGTCCCCACC  675
           |||||:|||::|  |||::::::::   :::
Db      9  ProAlaThrThrAlaThrAlaAlaGlu--SerValHisPheArgLeuCysProAla  27

Qy    676  CAGTTTGTTGGTGCCATCATCGGAAAGAGAGGGCTTGACCATAAAGAACATCACTAAGCAG  735
           ||||:::::  :::::::::
Db     28  ThrArgThrGlyAlaIleIleGlyLysArgGlyValIleGlySerValIleHisGlnSerVal  47

Qy    736  ACCCAGTCCCGGGTA  750
           |||||
Db     48  ThrGlySerLysIleArgValIleAspAspIleProValProSerGluArgValVal  67

Qy    751  ----------------------------GATATCCATAGAAAAGAGAAC  771
                                         :::::
Db     68  LeuIleIleAlaProSerGlyLysLysLysAspGluSerAsnValCysAspSerGluAsn  87

Qy    772  TCTGGAGCTGCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACT---------  822
           ||||:::
Db     88  ProGlySerGluGluProLysGlnLysGlyGlySerGluCysAlaGlyThrSerGlyGly  107

Qy    823  ----------------------------TCTGAAGCATGCCGGCCATGATGGTGAAATCATGCAGAAAA--  861
                                         :::|||
Db    108  AspAspGluGluAlaProSerSerAlaGlnMetAlaLeuLeuArgValPheGlnArgIle  127
```

```
US-09-270-437D-6 (1-3412) x T41600 (1-398)

Qy    505  GAGAACTACTCCTTCAAGATTCCTAC------ATCCCGGATGAAGAGTGAGCTCC---  555
           |||||||  ||||||  ::::::::::
Db     16  GluAsnSerPhePheAspValThrPheGlnAspGluProAspAspGluThrSerThrAla  35

Qy    556  -------------------------------CCTTCGCCCCCT---CAGCGAGGC  576
                                          ||||||
Db     36  ThrGlyIleAlaLysValSerIleProThrProLysProSerThrSerThrLeu  55

Qy    577  CAGCGTGGGACCACTCTTCCCGGAGCAAGGCCACGCCCCTGGGGGCACTTCTCAGGCC  636
           :::   |||  |||||||
Db     56  ThrArgAsnGlySerThrIleGlnGlnSerMetThrAsnGlnProGluProThrSerGlnVal  75

Qy    637  AGACAGATT-------------------------------------GATTTCCG  654
           |||
Db     76  ProProIleSerAlaLysProProMetAspAspAlaThrTyrAlaThrGlnGlnLeuThr  95

Qy    655  CTGCCGGATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGAGAGGGCTTGACC  714
           |||||  |||:::  |||||  ||||||||||||
Db     96  LeuArgAlaLeuLeuSerThrArgGluAlaIleGlyLysArgGlyAlaGlyLysLysAsn  115

Qy    715  ATAAAGAACATCACTAAGCAGACAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCT  774
           ::::
Db    116  ValAlaGluLeuArgSerThrArgAsnValLysGlyVal--ThrLysAlaValPro  134

Qy    775  GGAGCTGCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGC  834
           :::::::  |||:::  :::
Db    135  AsnValHisAspArgValLeuThrIleSerGlyLeuGluGluAsnValAlaArgAlaTyr  154

Qy    835  CGCATGATCCTTGAAATCATGCAGAAAGAGCA-------GATGAGCAAAACTAGCC  885
           ||||:::|||:::  :::
Db    155  ArgPheIleIleAspIlePheAlaLysLysAsnSerThrAsnProAspGlyThrProSerAsp  174

Qy    886  GAAGAATTCCT------CTGAAAATCTTGGCACCAATGGCTTGGTTGGAAGACTGATT  939
           ||||||||  :::::: ||||:::|||  ||||:::
Db    175  AlaAsnThrProArgGlyLeuArgLeuLeuIleAlaHisSerLeuMetGlySerIleIle  194

Qy    940  GGAAAAGAAGGCAGAAATTGAAGAAATTGAACATGAAACAGGGACCAAGATAACAATC  999
           ||||:::|||  :::  :::
Db    195  GlyArgAsnGlyLeuArgIleLysLeuLeuIleGlnAsnProLysCysSerCysArgMetIleAla  214

Qy   1000  TCATCTTTGCAGGATTGAGCATATACAACCCCGGAAAGAAGAACCATCACTGTGAAGGGCACA  1059
           :::||:::  :::::::::
Db    215  Ser-------LysAspMetLeuProGlnSerProArgThrValGluIleHisGlyThr  232

Qy   1060  GTTGAGGGCCTGTGCCAGTGCTGAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAA  1119
           ||||||||  :::  :::
Db    233  ValAspAsnLeuHisAlaAlaIleAlaIleTrpGluIleGlyLysCysSerLeuIleAspAspTrpGlu  252

Qy   1120  AATGATATGCTGGCTGTTAACCAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCCCA  1179
           ||||||
Db    253  Arg-------------------------GlyAla  255

Qy   1180  CTTGGCATCTTTCAACGACGTGTCCGTGCTATCCACCA----------------  1221
           ||||:::  :::|||
Db    256  GlyThrValPheTyrAsnProValSerArgLeuThrGlnProLeuProSerLeuAlaSer  275

Qy   1222  --GCAGGGCCCCGCGGA--GCTCCCCCGCTGCCCCCCCTTCACTACCCAC  1275
           |||||  :::::|||
Db    276  ThrAlaSerProGlnProGlnSerProProAlaAlaPro------SerThrThr  291

Qy   1276  TCCGGA-----------------TACTTCTCCAGCCTGTACCCCCATCAC  1308
           ||||||
Db    292  SerGlyGluAlaIleProGluAsnPheValSerTyrGlyAlaGlnValPheProAlaThr  311

Qy   1309  CAGTTTGGCCGTTGCCCGCATCACTCTTATCCAGACAGCAGAGATTGTGAATCTTTC  1368
           ||||:::  :::::::
Db    312  GlnMet---ProPheLeuGlnProLysValThrGln------AsnIleSer  326

Qy   1369  ATCCCAACCCAGGCTGTGGGCGCCATCATCGGAGAAGAGGGGGCCACACAAACAGCTG  1428
           ||||||  |||:::  :::::
Db    327  IleProAlaAspMetValGly...
```

```
Qy   1837  CCTCAGGGAGTCGCCTCACAGCCAGCCAGCAAGTGAGGCTCCCACAGGCACCAGCAAAACAAC 1896
                ||| ||   |||||     ||                     |||
Db    438  Pro--Gly-----ArgLeuProPheGlnGlyMetGlyProProProPheMetGly-        455

Qy   1897  GGATGAATGTAGCCCTTCAACACCTGACAGAATGAGACCAAACGCAGCCAGCAGATCG 1956
                ||||||||                         ::::
Db    456  ----ProTyrProGluProProProPheGlyPro--                      466

Qy   1957  GGAGCAAACCAAAGACCATCTGAGGAATGAGAAGTCTGCGGACGCGGCCAGGGACTCTGC 2016
                                                        |||||
Db    467  ------------------------------ArgGlnTyrProAlaSerProA          474

Qy   2017  CGAGGCCCTGAGACCCCAGGGCCG-------AGGAGGGGCGGGAAGGTCAGCCAGGTT 2070
                ||||                           ||||
Db    474  spArgTyrHisSerProValGlyProPheHisGluArgHisCysHisGly--ProGlyP 493

Qy   2071  TGCCAGAACCACCGAGCCCCGCC---TCCCGCCCCCAGGGCTTCTGCAGGGCTTCAGCCA 2127
                |||||        |||||
Db    493  heAspArgProGlyProGlyPheAspArgProPro--                       505

Qy   2128  TCCACTTCACCATCCACTCGGATCTCTCCTGAACTCCC 2165
                |||||
Db    506  -------------------SerProMetSerTrpThrPro 512
```

RESULT 5
T25832
hypothetical protein M01A10.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T25832
R;Scheet, P.
submitted to the EMBL Data Library, February 1997
A;Description: The sequence of C. elegans cosmid M01A10.
A;Reference number: Z20094
A;Accession: T25832
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;residues: 1-680 <SCH>
A;Cross-references: EMBL:U88174; PIDN:AAB42272.1; GSPDB:GN00019; CESP:M01A10.1
A;Experimental source: strain Bristol N2; clone M01A10
C;Genetics:
A;Gene: CESP:M01A10.1
A;Map position: 1
A;Introns: 63/2; 97/3; 121/1; 160/1; 269/3; 411/3; 512/1; 649/2

Alignment Scores:
Pred. No.:   3.4e-10   Length: 680
Score: 265.50   Matches: 118
Percent Similarity: 38.30%   Conservative: 103
Best Local Similarity: 20.45%   Mismatches: 213
Query Match: 4.34%   Indels: 143
Db: 2   Gaps: 17

US-09-270-437D-6 (1-3412) x T25832 (1-680)

```
Qy    424  GAAACCGCCGGTTGTCAACCTCACATATGCAACACAGAGAAGAGCAAAATAGCCATGGAG 483
                :::  ||||                       ||||      ::::
Db     60  AspLeuAlaAsnValAsnLysValLysTrpThrArgLeuPheGlnValLeuSerPheLys 79

Qy    484  AAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCAAGATTGA 543
                |||||||||            |||||
Db     80  AsnAsnPheSerHisPheAsnLeuArgHisPheLysIleLysAsnTyr-----Lys 97

Qy    544  GAGGTGAGCTCCCCTTCGCCCCTCAGCGAGCC---------CAGCGTGGGGACCAC 591
                |||||         ||||        ||||||
Db     98  HisIleSerSerThrAspThrAlaArgLysArgAspPheAlaArgSerGluGly 117

Qy    592  TCTTCCGGAGCAAGGCCACGCCCCTGGGGACACTTCTCAGGCCAGCAGATTGATTTC 651
                |||||||         |||||
Db    118  SerAspGluTyrGluGluTyrAlaProProCysLeuThrLysGlyAspIleAspTyr 137

Qy    652  CGGCTG-------------------------------CGGGATCCTGGTCCCCACCCAGTTGTT 684
                :::        :::  :::|||||
Db    420  ThrValGlnAspAlaLeuPheGlnIleLeuCysArgLeuArg----GluAlaMetPhe 437
```

```
Qy    862  -----------------------------GAGGCAGATGAGACCAAACTAGCCGAAGAGATT 894
                                             :::  ||||
Db    128  ValPheGlyAspAspAlaAlaThrValAspGlyAspGlyLeuAspGlyLysGluSerGlu 147

Qy    895  CCTCTG---AAAATCTTGGCACACAATGGCTTGGTTGGTGGAGACTGATTGGAAAAGAAGC 951
                ||||||          ::::                  |||||
Db    148  GlyLeuCysArgMetIleValArgGlyAsnGlnAsnValAspTyrLeuMetSerLysGlyGly 167

Qy    952  AGAAATTGAAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCCATCTTGCAG 1011
                :::  ||||||||||||                      ::::::||||
Db    168  LysMetIleGlnLysGlyLysIleArgGluAspSerGlyAlaIleValArgIleSerSerThrAsp 187

Qy   1012  GATTTG------AGCATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAG 1065
                 :::              |||                     :::  |||
Db    188  GlnIleProProCysAlaPheProGlyAspValValIleGlnMetAsnGlyLysPheSer 207

Qy   1066  GCCTGTGCCAGTGCTGAGATATAGAGAATTATGAAGAAGTCGCTTGAGGGCCTTTGAAAATGAT 1125
                |||||||          |||     |||
Db    208  SerValLysLysAlaLeuLeuLeuValThrAsnCysLeuGlnGluSer-----------    223

Qy   1126  ATGCTGGCTGTTAACCAACAGCCAATCTGATCCCCAGGGTTGAACCTGCGCACTTGGC 1185
                                                 |||
Db    223  -----------------------------------------------------      223

Qy   1186  ATCTTTTCAACAGGACTGTCCGTCGTATCTCCACCAGCAGGGCCCCGGGAGTCCCCC 1245
                                                            |||
Db    224  -----------------------------------------GlyAlaProPro      227

Qy   1246  GCT-----------------------GCCCCCTACCACCCCTTCACTACCCACTCCGGATAC 1284
                |||                       :::  :::
Db    228  ThrTrpAspGluCysProPheProGlnProGlyTyrProProGluTyrHisSerMetGlu 247

Qy   1285  TTCTCCAGCCTGTACCCCCATCACCAGTTTGGCCCGTTCCCGCAT--------------- 1329
                 :::  :::                     |||
Db    248  TyrHisProGlnTrpAspHisProProAsnProMetProGluAspValGlyProPhe 267

Qy   1330  CATCACTCTTATCCAGAGCAGGAGATTGTG--AATCTTTCATCCCAACCCAGGCTGTG 1386
                                              :::  |||
Db    268  AsnArgProValValGluGluValGlyGlyGlyAlaAlaPheArgLeuLeuCysProAlaAspLysLysVal 287

Qy   1387  GGGGCCATCATCGGGAGGAAGAGGGGGCACACATCAAACAGCTGGGCCGAGATTCGCCGGAGCC 1446
                |||||||||||       ||||              :::|||
Db    288  GlySerLeuIleGlyLysGlyLysGlyAlaGlyAlaValArgAlaLeuArgAlaAsnGluSerGlyAla 307

Qy   1447  TCTATCAAGATTGCC---CCTGCCGAAGGCCCAGAGCGTCAGCGAAAGAATGGTCATCATC 1503
                |||||||||          |||            ||::::::||||
Db    308  SerIleLysValSerAspProThrHis------AspSerGluGluArgIleIleValIle 325

Qy   1504  ACCGGGACCACCGGAAGCCCAGTTCAAG--------GCCCAGGGA------CGGATC 1545
                 :::  |||                     ||||
Db    326  SerAlaArgGluAsnLeuGluArgHisSerLeuAlaArgHisAspGlyValMetArgVal 345

Qy   1546  TTTGGAAACTGAAAGAGAGAAAACTCTTTAACCCCAAGAAGAAGTCGAAGCG 1605
                :::                     :::
Db    346  -----------------------GluProSerAlaAlaValAlaValAla 362

Qy   1606  CATATCAGATGCCCCTCTTCCACAGCTGGCCGGGTGATTGGCAAAGGTGGCAAGACCGTG 1665
                |||||              :::  :::
Db    363  ArgLeuLeuValHisSerProTyrIleGlyArgLeuLeuGlyLysLysGlyLysHisLeuIle 382

Qy   1666  AACGAGACTGCAGAACTTAACCAGTGCAGAAGTC--ATCGTGCCTCGTGCAGACAACGCCA 1722
                 :::|||||||||           :::  |||
Db    383  SerGluMetArgArgAlaThrGlyAlaSerIleArgValPheAlaLysAspGlnAlaThr 402

Qy   1723  -----GATGAAAATGAGGAGAGTGATCGTCAGAATTATCCGGCCACTTCTTGCTAGCCAG 1776
                    :::                       |||
Db    403  LysTyrGluSerGlnHisAspGluIleValIleGlyAsnLeu------Lys 419

Qy   1777  ACTGCACAGCGCAAGATGAAATTGTACAACAGGTGAAGCAGCAGGAGCAGAAATAC 1836
                 :::                         |||
Db    420  ThrValGlnAspAlaLeuPheGlnIleLeuCysArgLeuArg------GluAlaMetPhe 437
```

```
Qy   1705 CCTCGTGAC--------------------------------CAAACG 1719
          ||||||                                    ::|||
Db    447 SerArgAspProSerIleSerAlaIleGluLysGlnPheValIleArgGlySerGluThr 466

Qy   1720 CCAGATGAAAATGAGGAAGTGATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACT 1779
          |||::: |||::: ::|||:::|||||
Db    467 GlnValGluHisAlaLysHisLeuIleArgVal------------------------- 477

Qy   1780 GCACAGCGCAAGATCAGGGAAATT------------------------GTACAA 1809
          |||::|   ||| ||||||
Db    478 -------LysValGlyAspIleProProAsnThrProTyrIleAsnThrArgAlaAla 494

Qy   1810 CAGGTGAAGCAGCAGGAGCAGAAATACCTCAGGGAGTCGCCTCACAGCCGCAGCAAGTGA 1869
          |||::::     ::: |||
Db    495 GlnProLeuGlnPheSerHisGlnAsnProThrAlaIleAspSerTrpArgAlaGlnPro 514

Qy   1870 GGCTCCCACAGGCACCAGCAAACAACACGGATGAATGTAGCCCTTCCAACACCTGACAGAA 1929
          :::    |||:::|||
Db    515 PheThrThrGlnHisGlnAsnSerLeuSerLeuProGlnProGlnAlaHisGlnPhePro 534

Qy   1930 TGAGACCAAACGACCAGCCAGCCACGATCGGGAGCAAACCAAAGACCATCTGAG 1980
          |||||||                   :::
Db    535 AsnLeuMetAlaTyrSerAlaArgLeuGlyTyrGlnSerHisProProGln 551


RESULT 6
T27855
hypothetical protein ZK418.9 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T27855
R;Fulton, L.
submitted to the EMBL Data Library, April 1994
A;Description: The sequence of C. elegans cosmid ZK418.
A;Reference number: Z20430
A;Accession: T27855
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-768 <FUL>
A;Cross-references: EMBL:U00047; PIDN:AAA50693.1; CESP:ZK418.9
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:ZK418.9
A;Introns: 166/1; 221/3; 243/1; 279/1; 387/3; 490/1; 567/3; 727/2

Alignment Scores:
Pred. No.:                              Length:               768
Score:                    9.86e-10      Matches:              134
Percent Similarity:         259.00      Conservative:          86
Best Local Similarity:      38.33%      Mismatches:           218
Query Match:                23.34%      Indels:               136
DB:                          4.23%      Gaps:                  23
                                 2

US-09-270-437D-6 (1-3412) x T27855 (1-768)

Qy    560 CGCCCCTC---AGGGAGCCCAGCGTGGGGACCACTCTTCCCGGGAGCAGGCCACGCCC 616
          |||||           |||::: |||
Db    234 ArgProLeuAspSerGluIleLeuLeuAspGlyAspLeuIleProThrLysSerGlu 253

Qy    617 CTGGGGGCACTT---CTCAGGCGCAGACAGATTGATTTCCCGCTGCGGATCCTGGTCCCCA 673
          ::|||         ::: ||| |||
Db    254 ValGlyAspLeuAsnMet-GlyAspSerAspLysIleThrAspIleTyrProValProG 273

Qy    674 CCCAGTTGTTGGTGCCATCATCGGAAAGAGGGTTGACCATAAAGACCATCACTAAGC 733
          :::|||                   |||
Db    273 luLysValValGlyLeuValIleGlyLysGlyGlySerGluIleLeuArgGluGlnIle 293

Qy    734 AGACCAGTCCCGGGTAGATATCCATAGAAAAGAGACTCTGGAGCTGCAGAGAGCCTG 793
          :::
Db    293 hrSerGlyCysArgValGlnMetAspProAspHisGlnSerValAsnGlyPheAsnC 313

Qy    794 TCACCATCCATGCCACCCCAGAGGGACTTCTGAAGCATGCCGCATGATTCTTGAAATCA 853
          |||||                |||:::
Db    427 GlyArgGlyGlyGlyThrIleArgGlnIleLeuAsnLysGluSerGlyAlaTyrCysGluMet 446
```

```
Db    138 ArgValAspThrSerThrValIleLysAlaSerValSerIleProGluSerVal 157

Qy    685 GGTGCCATCATCGGAAAAGGAGGGGCTTGACCATAAAGAACATCACTAAGCCAGAGCCCAGTCC 744
          ::: |||||||        ||:::
Db    158 GlyLeuValIleGlyArgAsnGlyValGluIleGlnIleLeuSerGlnLysSerGlyCys 177

Qy    745 CGGGTAGATATCCATAGAAAAGAACTCGGAGCTGCAGAGAAGCCTGTCACCATCCAT 804
          ::: ||| |||
Db    178 ArgValGlnIle---ValAlaGluProSerThrThrGlyTyrArgSerValAspIleTyr 196

Qy    805 GCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATGATCTTGAAATCATGCAGAAAGAG 864
          |||      |||:::||
Db    197 GlyIleSerGluAsnIleGluValAlaLysLysLeuIleAsnGluValVal---- 213

Qy    865 GCAGATGAGACCAAACTAGCCGAAGAG--------ATTCCT 897
          |||      ::::||
Db    214 AlaArgGlyArgGlyLeuSerGlnIleGluProLeuProCysSerValProGlnPheGlnPro 233

Qy    898 CTGAAAATCTTGGCACACAATGGCTGGTT 927
          ::: |||:::::::|||
Db    234 IleProAlaValSerAsnSerLysValThrIleIleIleProIleAlaProAlaAsnLys 253

Qy    928 --GGAAGACTGATTGGAAAAGAAGGCAGAAATTGAAGAAATTGAACATGAAACAGGG 984
          ||     ::: ||:::|||
Db    254 CysGlyAlaIleIleGlyLysLysGlyGluGlnMetArgLysLeuArgSerTrpThrAsn 273

Qy    985 ACCAAGATAACAATCTCATCTTTGCAGAGATTTGAGCATTATACAACCCGGAAAGAACCATC 1044
          |||      :::        |||||
Db    274 CysAspPheIleLeu------IleGluGluAsnAsnIleAlaAspSerValLysProLeu 291

Qy   1045 ACTGTGAAGGGCACAGTTCAGCCCTTGCCAGTGCTGAGATAGAGATTATGAGAAGAGCTG 1104
          ::: ||| ||||
Db    292 GlnIleThrGlyGlnProLys--------------GluValGluHisAlaLysAlaLeu 306

Qy   1105 CGTGAGGCCTTTGAAAATGATATGCTGGCTGTTAACCAACAGCCAATCTGATCCCAGGG 1164
          |||     :::|||
Db    307 --------------------ValAlaAspIleLeuAspGly 313

Qy   1165 TTGAACCTGAGCCACTTGGCATCTTTCAACAGGACTGTCGTGCTATCTCCACCAGCA 1224
          :::|||||||
Db    314 PheAsp------------------------GluCysProProAla 320

Qy   1225 GGGCCCCGGCGAGCTCCCCGCTGCCCCGCCCTTCACTACCCACTCCGGATAC-------- 1284
          |||||| |||
Db    321 GlyMetAlaGlyAsnSerProValAlaAlaMetSer---------- 332

Qy   1285 TTCTTCCAGCCTGTACCCCCATCACCAGTTGGCCGTTCCCGCATCATCACTCTTATCCA 1344
          ::: ||| |||
Db    332 ---------------------------------------- 332

Qy   1345 GAGCAGGAGAGTTGTGAATCTTCTCATCCCCAACCCAGGCGTGTGGGCGCCATCATCGGGAAG 1404
          ||||||   :::|||
Db    333 ------------LeuGlnValLysValProArgSerThrValGlyAlaIleIleMetGlyLeu 348

Qy   1405 AAGGGGGACACATCAAACAGCTGGCGAGATTCGCCGGAGCCTCTATCAACAGATTGCCCCT 1464
          :::|||||||
Db    349 GlnGlySerAsnIleLysLysSerAsnGluThrArgLysIleIleGlnPheMetPro 368

Qy   1465 GCCGAAGGCCCAGACGTCAGCGAAAGATGGTCATCATCACCCGGCCACCGGAAGCCCAG 1524
          ::: ||| ::: |||
Db    369 AspAspAspProLysLeuMetGluArgThrLeuValLeuValIleGlyAsnLysVal 388

Qy   1525 TTCAAGGCCCAGGGACGGACGGATCTTTGGGAAACTGAAAGAGGAAAACTTCTTTAACCCCAAA 1584
          ::: |||:::
Db    389 TyrValCys--AlaArgLeuGlnLysIleValGlGluAlaAsnSerGluAsnAlaAsn 407

Qy   1585 GAAGAAGTGAAGCTGGAAGCCGCATATCAGAGTGCCCATCCTCCACCAGCTGGCCGGGTGATT 1644
          :::|||||
Db    408 ThrProIleSerLeu--PheTyrMetLeuIleProAlaSerLysSerCysGlyLeuValIle 426

Qy   1645 GGCAAAGGTGGCAAGACCGTGAACGAACTGACAGAACTTAACCAGTGACAGAAGTCATCGTG 1704
          |||||||:::||:::|||     :::
Db    427 GlyArgGlyGlyGluThrIleArgGlnIleLeuAsnLysGluSerGlyAlaTyrCysGluMet 446
```

RESULT 7

T19216
hypothetical protein C12D8.1a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T19216
R;McMurray, A.
submitted to the EMBL Data Library, June 1996
A;Reference number: 219092
A;Accession: T19216
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-589 <WIL>
A;Cross-references: EMBL:Z73969; PIDN:CAA98232.1; GSPDB:GN00023; CESP:C12D8.1a
A;Experimental source: clone C12D8
C;Genetics:
A;Gene: CESP:C12D8.1a
A;Map position: 5
A;Introns: 7/3; 25/1; 159/3; 318/3; 513/2

Alignment Scores:
Pred. No.:   589
Score: 254.50   Length: 589
Percent Similarity: 38.31%   Matches: 118
Best Local Similarity: 22.61%   Conservative: 82
Query Match: 4.16%   Mismatches: 180
DB: 2   Indels: 142
                                                Gaps: 19

US-09-270-437D-6 (1-3412) x T19216 (1-589)

```
Db   134 ThrIleAspIleAlaIlePro------------------------ProAsnArgCysGlyLeuIle 147

Qy   937 ATTGGAAAAGAAGGCAGAAATTTGAAGAAAATTGAACATGAAACACAGGACCAAGATAACA 996
             |||||||    ::::|||||||    :::::|||     ||:::
Db   148 IleGlyLysSerGlyAspThrIleArgGlnLeuGlnLysSerGlyIleLysCysMetIle 167

Qy   997 ATCTCATCTTTGCAGGAT---TTGAGCATATACAACCCGGAAAGAACCATCACTGTGAAG 1053
             :::         ||||||    :::::||||||    ::: |||||||||||
Db   168 Leu------ValGlnAspAsnGlnSerValSerArgSerIleSerProLeuArgIleThr 185

Qy  1054 GGCACAGTTGAGGGCTGTGCCAGTGCTGCTGACAGTAGAATTATGAGAAGGTGCGTGAGGCC 1113
             ::|||||||    ||:::|||    :::
Db   186 GlyAspProGln--------------LysIleGluLeuAlaLysGlnLeu---------- 197

Qy  1114 TTTGAAAAATGATATCTGGCTGTTAACCAACAAGCCAATCTGATCCCAGGGTTGAACCTC 1173
                                                       ::: |||||
Db   198 -----------------------------------------ValAlaGluIleLeuAsnSer 204

Qy  1174 AGCGCACTTGGCATCTTTCAACAAGGACTGTCCGTGCTATCTCCACCAGCAGGGCCCCGC 1233
             :::|||||||    :: :::
Db   205 GlyGlyAspGlyAsnGlyGlySerGlyLeuGlnMet---------------------- 216

Qy  1234 GGAGCTCCCCCGCTGCCCCCTACACCCCTTCACTACCCACTCCGGATACTTCTCCAGC 1293
                                                                       
Db   216 -------------------------------------------------------- 216

Qy  1294 CTGTACCCCCCATACCACCAGTTTGGCCCGTTCCCGCATCATCACTCTTATCCAGAGCAGGAG 1353
             :::                                         :::
Db   217 -------HisHisAlaGlyGly--------GlyGlyValAlaSerAlaArgGlyGlu 230

Qy  1354 ATTGTGAATCTCTTCATCCCAACCCAGGCTGTGGCGGCCCATCATCCGGAAGAAGAGGGGCA 1413
             ::|||    :::|||||||||||||    ::||||||||||
Db   231 ValVal-----------ValProArgSerValGlyIleIleIleGlyGlnGlyAspAsp 247

Qy  1414 CACATCAAACAGCTCGGCGAGATTGCGCGAGCCCTATCAAGATTGCCCCTGCGGAAGGC 1473
             |||||||    ::::|||||||    :::
Db   248 ThrIleLysArgLeuAlaMetGluThrGlyThrLysIleGlnPheLysProAspAspAsp 267

Qy  1474 CCGACGTCAGCGAAAGGATGGTCATCATCACCGGGCCACCGAAGCCCAGTTCAACGCC 1533
             |||||||    :::::|||||||    ::::::
Db   268 ProSerThrProGluArgCysAlaValIleMetGlyThrArgAspGlnIleTyrArgArgAla 287

Qy  1534 CAGGGACACGGATCTTTGGGAAACTGAAAGAGGAAAACTTCTTTAAC---------- 1578
             ||||||||||||    :::::
Db   288 ThrGluArgIleThrGluLeuValLysLysSerThrMetGlnGlnGlyGlyGlyGlyAsn 307

Qy  1579 ---CCCAAAGAAGAAGTGAAGCTGGAAGCG--------CATATCAGAGTGCCCCTCTTCC 1626
                |||||    :::::||||||    ::::::|||
Db   308 ValAlaGlyValAlaMetValSerAsnGluAlaSerThrPheTyrMetSerValProAlaAla 327

Qy  1627 ACAGCTGGCCGGTGATTGGCAAAGGTGCCAAAGGTGGAACGTGAACGAACTGCAGAACTTAACC 1686
             ||||||||    :::::|||||||    ::||||||
Db   328 LysCysGlyAsnAlaIleLysGlyLeuValIleGluLysGlnIleLeuAsnSerGluSer 347

Qy  1687 AGTGCAGAAGTGCATCGTCGTGACCAAACGCCAGATGAAAATGAGGAAGTGATCGTC 1746
             ||||||    :: :::::|||    ::||||||||||
Db   348 GlyValAlaHisCysGlyLeuSerGluArgAspProThrGlyAsnAlaAspGluLysValPheVal 367

Qy  1747 -----------AGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAG 1791
                          :::||||||    ::||||||
Db   368 IleLysGlyLysArgGlyIleGluHis------AlaLysHisLeuIleArgGlyIleLys 385

Qy  1792 ATCAGGGGAAATTGTACACAGGTGAAGCGAGCAGGAGCAGATAAATACCCTCAGGGAGTGCC 1851
             ::: :::|||    ||||||    :::
Db   386 ValGlyValAspIleAlaProAsnThr------------------------ProAsnArgCysGlyLeuLeuIle 393

Qy  1852 TCACAGCGCAGCAAGTGAGGGCTCCCACAGCACCAGCAAAACAACGGATGAATGTAGCCC 1911
                                                                :::|||
Db   394 -----------------------------------------------------Pro 394

Qy  1912 ATTGGAAAAGAAGGCAGAAATTTGAAGAAAATTGAACATGAAACACAGG------CCAGCCAGA 1953
             |||||||    ::||||||    ::|||
Db   395 PheThrGlyAspAspSerAlaMetThrMet------GlnThrGlnPheSerAlaProAlaGln 412

Qy  1954 TCG------GGAGCAAACCAAAGACCATCTGGAGGAATGAGAGAGTCTGCGGAGGCGGCCAG 2007
             :::          |||    :::  ::
Db   413 AsnAsnPheGlyGlyGlnGlyGlnLeuGlnTrpAsnProValAlaGlnIleProAlaAlaAlaGln 432

Qy  2008 GGACTCTGCCGAGGCCCTGAGACCCCAGGGCCGCGAGGGGCGGGAAGTCAGCGCAG 2067
                               ||||||    :: :::
Db   433 ---------------------------AsnProTyrArgLeuValGlyGlyTrpGlnGlnAsnSerVal 445

Qy  2068 GTTTGCCAGAACCACCACGAGCCCGCCCTCCCCGCCCCCCCCCAGGGCGTTCTGCAGGCTTCAGCCA 2127
             |||||    ||:::|||||||||
Db   446 TyrAlaGlnGlnThrAlaAlaProAlaAlaAlaProTyr-AlaAlaAlaGlyIleAlaValGl 465

Qy  2128 TCCA 2131
             |||
Db   465 nPro 466
```

RESULT 8

T19217

```
hypothetical protein C12D8.1b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T19217
R;McMurray, A.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19092
A;Accession: T19217
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-611 <WIL>
A;Cross-references: EMBL:Z73969; PIDN:CAA98233.1; GSPDB:GN00023; CESP:C12D8.1b
A;Experimental source: clone C12D8
C;Genetics:
A;Gene: CESP:C12D8.1b
A;Map position: 5
A;Introns: 181/3; 340/3; 535/2
```

Alignment Scores:

```
Pred. No.:                                                    611
Score:              1.89e-09        Length:                   611
                    254.50          Matches:                  118
Percent Similarity: 38.31%          Conservative:              82
Best Local Similarity: 22.61%       Mismatches:               180
Query Match:        4.16%           Indels:                   142
DB:                 2               Gaps:                      19
```

US-09-270-437D-6 (1-3412) x T19217 (1-611)

```
Qy   667 GTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAACATC 726
             |||||    :::::|||||||    :::::|||
Db    76 IleProGluSerAlaValGlyIleValIleGlyGlySerGlyLeuGlnIleGlnIleGlnIle 95

Qy   727 ACTAAGCAGACAGACCCAGTGCCCGGTAGATATCCATAGAAAAGAGAACTCTGGAGCTCCAGAG 786
             ||||||                            :::
Db    96 GlnAlaLysAlaGlyLysAlaGlyLysArgValGlnMetSerProAlaAspProSerGlyVal 115

Qy   787 AAGCCTGTCACCATCCATGCCACCCCAGAGGGACTTCTGAAGCATGCCGCATGATTCTT 846
             |||    :: :::|||
Db   116 ArgMetValThrLeuGluLysGlySerArgSerAsnValGluThrAlaLysHisLeuIleLeuAsn 135

Qy   847 GAAATCATCGAAGAAAGAGAGGCAGAT------------------GAGACC 876
             |||||    :::      ::
Db   136 GluValValAlaArgSerGlnAsnProArgProGlnTyrGlyPheProAlaGlnThr 155

Qy   877 AAACTAGCCGAAGACAGAGATTCCTCGAAAATCTTGGCACACATATGGTTGGTTGGAAGACTG 936
             :::                                         :::
Db   156 ThrIleAspIleAlaIlePro------------------------ProAsnArgCysGlyLeuIle 169

Qy   937 ATTGGAAAAGAAGGCAGAAATTTGAAGAAAATTGAACATGAAACACAGGACCAAGATAACA 996
             |||||||    ::::|||||||    :::::|||     ||:::
Db   170 IleGlyLysSerGlyAspThrIleArgGlnLeuGlnLysSerGlyIleLysCysMetIle 189

Qy   997 ATCTCATCTTTGCAGGAT--TTGAGCATATACAACCCGGAAAGAACCATCACTGTGAAG 1053
             :::         ||||||    :::::||||||    ::: |||||||||||
Db   190 Leu------ValGlnAspAsnGlnSerValSerArgSerIleSerProLeuArgIleThr 412
```

```
Qy    2068 GTTTGCCAGAAACACCGAGCCGCCTCCGCCTCCGCCCCCAGGGGTTCTGCAGGTTCAGCCA 2127
                ||||
Db     468 TyrAlaGlnThrAlaAlaAlaProAlaAlaAlaProTyr-AlaAlaAlaGlyIleValGl 487
                                                         ::::||||

Qy    2128 TCCA 2131
           ||||
Db     487 nPro 488
```

RESULT 9
A53184
myc far upstream element-binding protein - human
N;Alternate names: FUSB-binding protein
C;Species: Homo sapiens (man)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 05-Nov-1999
C;Accession: A53184
R;Duncan, R.; Bazar, L.; Michelotti, G.; Tomonaga, T.; Krutzsch, H.; Avigan, M.; Levens
Genes Dev. 8, 465-480, 1994
A;Title: A sequence-specific, single-strand binding protein activates the far upstream
A;Reference number: A53184; MUID:94170991; PMID:8125259
A;Accession: A53184
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-644 <DUN>
A;Cross-references: GB:U05040; NID:g460151; PIDN:AAA17976.1; PID:g460152
C;Keywords: DNA binding

```
Alignment Scores:
Pred. No.:                       1.01e-08      Length:            644
Score:                            244.00       Matches:           145
Percent Similarity:               35.64%       Conservative:       82
Best Local Similarity:            22.76%       Mismatches:        241
Query Match:                       3.99%       Indels:            169
DB:                                   2        Gaps:               27
```

US-09-270-437D-6 (1-3412) x A53184 (1-644)

```
Qy     336 CCCTCCTCACCTGCAGTGGGAGGTGTTGATGGACTTTTGGCTCAATATGGACCAGTGGA 395
               ||||||||              :::||             ||||||
Db      10 ProSerGlySerAlaGlyGlyGlyGly-----------------GlyGlyGly 22

Qy     396 GAATGTGAACAAGTCAACACAGACACAGAAACGCCGTTGTCAACGTCACATATGCAAC 455
               ||| ||||     :::               |||::  -------AlaAr 39
Db      23 GlyGlyGlyGly-GlyValAsnAspAlaPheLysAspAlaLeuGlnArg-------

Qy     456 AAGGAGAAGAAGCAAAAATAGCCATGGAGAAGAGCTAAGCGGCATCAGTTTGAGAACTACTC 515
               :::                   |||       ::||||          |||||
Db      39 gGlnIleAlaAlaLysIleGlyGlyAspAlaGlyThrSerLeuAsnSerAsnAspTyrGl 59

Qy     516 CTTCAAGATTTCCTACATCCGGATGAAGAGGTGAGCTCCCCTTCGCCCCCTCAGCGAGC 575
               :||                  ||||       |||::
Db      59 yTyrGlyGlyGlnIleLysArgProLeuGluAspGlyAspGlnLysLysValVal 79

Qy     576 CCAGCGTGGGAGACCACTCTTCCCGGAGCAAGGCCACCGCCCTGGGAGCCACTTCTCAGGC 635
                                                           :::
Db      79 aProGlnAsnAspSerPheGlyThrGlnLeu------ProProMetHisGlnGlnGlnSe 97

Qy     636 CAGACAGATGATTTCCCGTGCGATCCTGCCGGATCCTGCTCCACCCAGTTGTTGGTGCCATCAT 695
               :|||                 ::||        ||||
Db      97 rArgSerVal--MetThrGluGluTyrLysValProAspGlyMetValGlyPheIleIl 116

Qy     696 CGGAAAGGAGGGCTTGACCATAAAGACACATCACTAAGCAGACCCAGTCCCGGGTAGATAT 755
               ||||:::   |||          :::::::
Db     116 eGlyArgGlyGlyGluGlnIleSerArgIleGlnGlnSerGlyCysLysIleGlnIl 136

Qy     756 CCATAGAAAGAGAACTCTGAGCTGCAGAGAAGCCTGTCACCATCATGCCACCCCAGA 815
               |||||               ::::::
Db     136 e---AlaProAspSerGlyLeuProGluArgSerCysMetLeuThrGlyThrProGl 155

Qy     816 GGGGACTTCTGCAAGCATGCCGCATGATTCTTGAAATCATGCAGAAA------------ 861
               ||||                     ||||||:::::||
Db     155 uSerValGlnSerAlaLysArgLeuArgLeuGluValGlyArgProAlaPr 175
```

```
Db     190 Leu------ValGlnAspAsnGlnSerValSerAspGlnSerLysProLeuArgIleThr 207

Qy    1054 GGCACAGTTGAGGCCTGTGCCAGTGCTGAGATAGAGATTATGATGAAGAAGCTGCGTGAGGCC 1113
               |||             :::||||||||          :::|||||:::||
Db     208 GlyAspProGln-----------------LysIleGluLeuAlaLysGlnLeu----- 219

Qy    1114 TTTGAAAATGATATGCTGGCTGTTAACCAACAAGCCAATCTGATCCCAGGGTTGAACCTC 1173
                                                :::           ||||||
Db     220 ------------------------------ValAlaGluIleLeuAsnSer 226

Qy    1174 AGCGCACTTGGCATCTTTCAACAGGACTGTCCGTGCTATCTCCACCAGCAGGGCCCCGC 1233
               ||||||||   :::||||||    :::
Db     227 GlyGlyAspGlyAsnGlyGlySerGlyLeuGlnMet------------------- 238

Qy    1234 GGAGCTCCCCGCCGTGCCCCCTACCACCCCTTCACTACCCACTCCGGATACTTCTCCAGC 1293
               |||
Db     238 ------------------------------------------------- 238

Qy    1294 CTGTACCCCCATCACCAGTTGGCCGTTCCCGGATCATCACTCTTATCCAGAGCAGGAG 1353
               ||||||                     :::
Db     239 -------HisHisAlaGlyGly--------GlyGlyAlaSerAlaArgGlyGlu 252

Qy    1354 ATTGTGAATCTTCTTCATCCCAACCCAGGCTGTGGGGCGCCCATCATCGGGAGAAGGGGGCA 1413
               :::|               :::||||||   |||||||||:::
Db     253 ValVal--------ValProArgSerValGlyIleIleGlyLysGlnGlyAsp 269

Qy    1414 CACATCAAACAGCTGCGAGATTCGCGGACGCCTCTATCAAGATTGCCCCTGCCGAAGGC 1473
               ||||||:::||||::||||          |||   :::
Db     270 ThrIleLysArgLeuAlaMetGluThrGlyThrLysIleGlnPheLysProAspAspAsp 289

Qy    1474 CCAGACGTCAGCGAAAGGATGGTCATCATCACCGGGCCACCGGAGCCCAGTTCAAGGCC 1533
               |||||   :::           ||||||           :::||||||
Db     290 ProSerThrProGluArgCysAlaValIleMetGlyThrArgAspGlnIleTyrArgAla 309

Qy    1534 CAGGGACGGATCTTTGGGAAACTGAAAGAGGAAAACTTCTTAAC----------------- 1578
               ||||||||                ::::::
Db     310 ThrGluArgIleThrGlyLeuValLysLysSerThrMetGlnGlnGlnGlyGlyGlyAsn 329

Qy    1579 --CCCAAAGAAGAAGTGAAGCTGGAAGCG--------CATATCAGAGTGCCCTCTTCC 1626
               |||||||             :::::: ||||||  :::::::
Db     330 ValAlaGlyAlaMetValSerAsnGluAlaSerThrPheTyrMetSerValProAlaAla 349

Qy    1627 ACAGTTGGCCGGTGATTGGCAAAGTGGCAAGACCGTGAACGAACTGCAGAACTTAACC 1686
               :::||||        ::: :::         |||
Db     350 LysCysGlyLeuValIleGluGlyLysGlyGlyGlyValGluThrIleLysGlnIleLeuAsnSerGluSer 369

Qy    1687 AGTGCAGAAGTCATCGTGCCCGTGACCAAACGCCAGATGCAAAAATGAGGAAGTGATCGTC 1746
               ||||||            :::|||||||  |||    ::||||::
Db     370 GlyAlaHisCysGluLeuSerGluAspArgAspThrGlyAsnAlaAspGlyLysValPheVal 389

Qy    1747 -------------------AGAATTATCGGGCACTTCTTTGCTCAGCCAGACTGCACAGCCAG 1791
                                  ||||||||||                :::
Db     390 IleLysGlyLysLysGlyArgAlaIleGluHis-----AlaLysHisLeuIleLeuArgIleLys 407

Qy    1792 ATCAGGGAAATTGTACAACAGTGAAGCAGCAGAGAGAATACCCTCAGGGAGTCGCC 1851
               :::  ::||||
Db     408 ValGlyAspIleAlaProAsnThr------------------------------ 415

Qy    1852 TCACAGCGCAGCAAGTGAGGTCCCACAGGACACCAGCAAAACAACGGATGAATGTAGCCC 1911
               :||| :::                                        |||
Db     416 --------------------------------------------------Pro 416

Qy    1912 TTCCAACAC----------CTGACAGAATGAGAACAGCAAACGCAG-------CCAGCCAGA 1953
               ||||:::           |||||                 ::||||||
Db     417 PheArgAspAspSerAlaMetThrMet-----GlnThrGlnPheSerAlaProAlaGln 434

Qy    1954 TCG------GGAGCAAACCAAAGACCATCTGAGGAATGAGAAGTCTGCGGAGGCGGCCAG 2007
               :::       |||||||||      :::::::
Db     435 AsnAsnPheGlyGlyGlnInTrpAsnProValAlaGlnIleProAlaAlaAlaGln 454

Qy    2008 GGACTCTGCCGAGGCCCTGAGAACCCCAGGGCCGAGGGGGCCGGGGAAGGTCAGCCAG 2067
               ||||||                      :::
Db     455 ------------------AsnProTyrGlnValGlyTyrGlnInGlnAsnSerVal 467
```

Db   455 yProProValProHisGlyPrcHisGlyVal-ProGlyProHisGlyProProGlyProP 475
            ||| |||||||
Qy   1878 CAGGCACCAGCAAACAACAACGATGATGAATGGAGCCCTTCCAACACCTGACACACTGAGACCA 1937
            ||| || |||
Db   475 roGly-ProGly---------------------ThrPro------MetGlyPro 483
            ||| || |||
Qy   1938 AACGGAGCCAGCCAGATCGGGAGCAACAACCAAAGACCATCTGAGGAATGAGAAGTCTCGG 1997
                :::
Db   484 TyrAsnProAlaProGlyAsnProGlyPro---------------------GlyPr 493
            |||
Qy   1998 AGGGGGCCAGGGACTCTGCCGAGG---CCCTGAGAACCCCAGGGGCCGAGGAGGGGCGGG 2054
            ||| ||| |||:::
Db   494 ------ProGlyProAlaProHisGlyProProAlaProTyrAlaProGlnGlyTrpGly 511
            ||||| |||
Qy   2055 GAAGGTCAGCCAGGTTTGCCAGAA-----CCACCGAGCCCCGCC 2093
            |||||
Db   512 AsnAlaTyrProHisTrpGlnGlnGlnAlaProProAla 526

RESULT 10
S58529
alpha-complex protein 1 - human
N;Alternate names: nucleic acid-binding protein; protein PCBP-1
C;Species: Homo sapiens (man)
C;Date: 15-Feb-1996 #sequence revision 01-Mar-1996 #text change 01-Dec-2000
C;Accession: S58529; S65678; S43489; S41378; S42472
R;Kiledjian, M.; Wang, X.; Liebhaber, S.A.
EMBO J. 14, 435-4364, 1995
A;Title: Identification of two major cellular poly(rC)-binding human proteins in the alpha-globin mRNP stability com
A;Reference number: S58523; MUID:9616208; PMID:7556077
A;Accession: S58529
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-356 <KI1>
A;Accession: S58523
A;Status: preliminary
A;Molecule type: protein
A;Residues: 125-139;231-265;315-323 <KI2>
R;Leffers, H.; Dejgaard, K.; Celis, J.E.
Eur. J. Biochem. 230, 447-453, 1995
A;Title: Characterisation of two major cellular poly(rC)-binding human proteins, each co
A;Reference number: S65678; MUID:95331278; PMID:7607214
A;Accession: S65678
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-204,'V',206-356 <LEF>
A;Cross-references: EMBL:X78137; NID:g460770; PIDN:CAA55016.1; PID:g460771
A;Experimental source: AMA cells (transformed human amnion cells)
A;Note: submitted to the EMBL Data Library, March 1994
R;Aasheim, H.C.; Loukianova, T.; Deggerdal, A.; Smeland, E.B.
Nucleic Acids Res. 22, 959-964, 1994
A;Title: Tissue specific expression and cDNA structure of a human transcript encoding a
A;Reference number: S43489; MUID:94203810; PMID:8152927
A;Accession: S43489
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-298,'H', <AAS>
A;Cross-references: EMBL:Z29505; NID:g444020; PIDN:CAA82631.1; PID:g444021
C;Genetics:
A;Gene: GDB:HNRNPX
A;Cross-references: GDB:344947
C;Keywords: RNA binding

Alignment Scores:
Pred. No.:   2.06e-08   Length: 356
Score: 238.50   Matches: 89
Percent Similarity: 41.78%   Conservative: 66
Best Local Similarity: 23.99%   Mismatches: 137
Query Match: 3.90%   Indels: 79
DB: 2   Gaps: 13

US-09-270-437d-6 (1-3412) x S58529 (1-356)

Qy   862 --------GAGGCGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAATCTTGGC 911
            ||| :::|||
Db   175 oGlyPheHisHisGlyProGlyProGlyAsnAlaValGln------GluIleMet1 192
            ||| ::|||||
Qy   912 ACACAATGGCTTGGTTGGAGACTGATTGGAGAAAGAGGCAGAAATTGAAGAAATTGA 971
            :::|||| |||
Db   192 eProAlaSerLysAlaGlyLeuValIleGlyLysValGlyGluThrIleLysGlnLeuGl 212
                    |||
Qy   972 ACATGAAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTGAGCATATACACCC 1031
            :::|||||| |||
Db   212 nGluargAlaGlyValValsMetValMet------IleGlnAsp 226
                    |||
Qy   1032 GGAAAGAAACCATCACTGTGAAGGGACAGTTGAGCCTGTGCCAGTGCTGAGATAGAGAT 1091
            |::|| |||:::
Db   226 oGlnAsnThr----------------------GlyAlaAsp-------- 232
            |::||
Qy   1092 TATGAGAAGCTGCGTGAGGCCTTGAAAATGATATGCTGGCTGTTAACCAACAGCCAA 1151
            |||||| |||
Db   233 ---LysProLeuArg------IleThrGlyAspProTyrLysValGlnGlnAlaLysGl 249
            |: ::: |||
Qy   1152 TCTGATCCCCAGGGTTGAACCTCAGCGCACTTGCATCTTTTCAACAGGACTGTCCGTGCT 1211
            ::::: ||| :::::
Db   249 uMetVal---------------------------LeuGluLeuI 255
            :::::
Qy   1212 ATCTCCACCAGCACCAGGGCCCCCGGAGCTCCCCCGCCGCCCCCTACCCCCTTCACTAC 1271
            : |||
Db   255 eArgAspGInGlyGlyPheArgGluVal------------------- 264
                    |||
Qy   1272 CCACTCCGGATACTTCTCCCAGCCTGTACCCCCATCACCACCAGTTTGGCCCGTTCCCGCATCA 1331
            :::::::||| |||
Db   265 ----------ArgAsnGluTyrGly--------- 269
                    |||
Qy   1332 TCACTCTTATCCAGAGCAGGAGATTGTGAATCTCTTCATCCCAACCCAGGCTGTGGGCGC 1391
            ||| |||
Db   270 -SerArgIleGlyGlyAsnGluGlyIleAspValProIleProArgPheAlaValGlyIl 289
            |||
Qy   1392 CATCATCGGGAAGAAGAGGGGGCCACACATCAAACAGCAGACAGTCTGGCGGAGCCTCTAT 1451
            :::||||||| ||||||
Db   289 eValIleGlyArgAsnGlyGluMetIleLelysGlySileGlnAsnAspAlaGlyValArgIl 309
            :::|||||||
Qy   1452 CAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGGAAAGGATGGTCATCACCGGGCC 1511
            |:::|| |||
Db   309 eGlnPheLysProAspAspGly---ThrThrProGluArgIleAlaGlnIleThrGlyPr 328
            |:::||
Qy   1512 ACCGGAA----GCCCAGTTCAAGGCCCAG 1536
            ||||:::
Db   328 oProAspArgCysGlnHisAlaAlaGluIleIleThrAspLeuLeuArgSerValGlnAl 348
            ||||:::
Qy   1537 ----------GGACGATCTTTGGGAAA 1554
            |||||
Db   348 aGlyAsnProGlyProGlyGlyArgGlyArgGlyGlyGlnGlyAsnTr 368
            |||||
Qy   1555 -----------CTGAAAGAGGAGAAACTTCTTTAACCCCAAAGAGAAGT 1592
            |||
Db   368 pAsnMetGlyProGlyLeuGlnGluPheAsnPheIle-- 382
            |||
Qy   1593 GAAGCTGGAAGCGCCATATCCAGAGCGCCCCTTCCACAGTGCCGGGTGATTGGCAAAGG 1652
            ||||| :::|||||
Db   383 ----------ValProThrGlyLysThrGlyLeuIleGluGlyLysGl 395
            |||||
Qy   1653 TGGCAAGACCGTGAACGAACTGCAGAACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGA 1712
            ::: ||| :::
Db   395 yGlyGluThrIleLysSerIleSerGlnAsnLeuArgIleAlaArgIleGluGlnnArgAs 415
            :::
Qy   1713 CCAAACGCCAGATGAAAATGAGGAAGTG--ATCGTCAGAATTATCGGGCAC--------- 1761
            ||| ::::::||| |||
Db   415 nProProAsnAlaAspProAsnMetLysPheThrIleArgGlyThrProGlnGl 435
            ||| :::::
Qy   1762 ---TTCTTTGCTAGCCAGACTGCCACAGCGCAAGATCACAGGGAAATTGTACACAGGTGAA 1817
            ::: |||||
Db   435 nIleAspTyrAlaArgGlnLeuIleGluGluIleSerIleGluGlyIleProValAsnProLeuGl 455
            :::
Qy   1818 GCAGCAGGAGCAGAAATACCCTCAGGGAGTGCGCCTCACAGCGCCAGCAAGTGAGGCTCCCA 1877

Db   340 GlnTyrLeuIleAsnAlaArgLeuSerSerGlu 350
                 |||

RESULT 11
T04533
hypothetical protein F28J12.30 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 14-May-1999
C;Accession: T04533
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,
submitted to the Protein Sequence Database, February 1998
A;Reference number: 215377
A;Accession: T04533
A;Molecule type: DNA
A;Residues: 1-846 <BEV>
A;Cross-references: EMBL:AL021710
A;Experimental source: cultivar Columbia; BAC clone F28J12
C;Genetics:
A;Map position: 4
A;Introns: 94/3; 140/3; 161/3; 191/3; 232/3; 262/1; 287/1; 293/3; 300/3; 503/3; 675/3;
A;Note: F28U12.30

Alignment Scores:
Pred. No.:
Score: 3.59e-08   Length: 846
                                236.50   Matches: 122
Percent Similarity: 35.14%   Conservative: 92
Best Local Similarity: 20.03%   Mismatches: 229
Query Match: 3.86%   Indels: 166
DB: 2   Gaps: 19

US-09-270-437D-6 (1-3412) x T04533 (1-846)

Qy 370 CTTTTGGCTCAAATATGGGACACAGTGGAGAATGTGGAGAACAAGTCAACACA----------- 417
            ||| : :: :::
Db 268 LeuLeuAspSerTyrTyrGly------------HisThrIleGlyValAsnThrAlaThrPheThr 284

Qy 418 ---------------------------------GACACAGAAACCGCCGTTGTCAACGTC 444
                                             ::||| || :: :||| :::|||
Db 285 ArgLysGlySerGlyMetSerSerGlyLeuIleGluThrAspSerValLeuAsnSer 304

Qy 445 ACATATGCAACAAGAGAGAAGAAGCAAAATAGCAGGGGCATCAAGGGCATCAGTT 504
                              |||
Db 305 ValTyrLeuThrMetValGluArgLys------ 313

Qy 505 GAGAACTACTCCTTCAAGATTCCTACATCCCGGATGAAGAGGTGAGTCCCCTTCGCCC 564
Db 314 --------------------------Lys 314

Qy 565 CCTCAGCGAGCCCAGCGTGGGGACCACTCTTCCCGGGACCAAGGGGGGC 624
            :::
Db 315 ArgLysGlnIleGlnArgAsnAsnSerGluSerAsnArg------------ 327

Qy 625 ACTTCTCAGGCCAGCACCAGATTGATTTCCCG-----------------CTG 657
Db 328 ---AsnGlnIleLysArgArgIleSerHisAspLysIleAsnArgAspGluLeuValTyr 346

Qy 658 CGGATCCCTGGTCCCCACCCAGTTGTTGGTGCCCATCATCGGAAAGGAGGGCTTGACCATA 717
            |||||| :::|||||
Db 347 ArgIleLeuCysProIleAspValValGlyValIleGlyValIleGlyLysSerValIle 366

Qy 718 AAGAACATCACTAGCAGCAGACCCAGTCCCGGTAGATATCCATAGAAAGAGAACTCTGGA 777
            ||||||||| :::
Db 367 AsnAlaIleAlaArgHisAsnThrLysAlaLysIleLysValPhe---AspGlnLeuHisGly 385

Qy 778 GCTGCAGAGAAGCCTGTCACCATCCATGCCGAAGCATGCCGGC 837
            || :::
Db 386 CysSerGlnArgValIleThrIleTyrCysSerValLysGluLysGlnGlu------ 403

Qy 838 ATGATTCTTGAAAATCATGCAGAAAGAGGGCAGAATGAGACCCAAGAGATTCCT 897
            :::   :::
Db 404 ------IleGlyPheThrLysSerGluAsnGluProLeuCysCysAlaGlnAspAlaLeu 421

Qy 898 CTGAAAATCTTGGCACACAATGGCTTGGTT------ 927

Qy 619 GGGGGCACTTCTCAGGCCAGACAGATTGATTCCCGTGCGGATCCTGGTCCCACCCAG 678
            ::||| :::|||
Db   4 GlyValThrGluSerGlyLeuAsnValThrLeuThrIleArgLeuLeuHisGlyLys 23

Qy 679 TTTGTTGGTGGCCATCATCGGAAAGGAGGGCTTGACCATAAAGACACATCACTAAGCAGACC 738
            |||||| ::::::||||| :::|||
Db   24 GluValGlySerIleIleGlyLysLysGlyGluSerValLysArgIleArgGluGluSer 43

Qy 739 CAGTCCCGGGTAGATATCCATAGAAAGAGAACTCTGGAGCTGCAGAGAAGCCTGTCACC 798
            :::||||::::::||| ||:::
Db   44 GlyAlaArgIleAsnIle-------SerGluGlyAsnCysProGluArgIleIleThr 60

Qy 799 ATCCATGCCGAAGCATGCCGGCGGGACTCTGAAGCATGCCGATGATTCTTGAAAATCATGCAG 858
            :::
Db   61 LeuThrGlyProThrAsnAlaIlePheLysAlaPheAlaMetIleIleAspLysLysLeuGlu 80

Qy 859 AAAGAGGCAGATGAGACCAAACTA----------GCCGAAGAGATTCCTCTG 900
            :::   :::   :::|||
Db   81 GluAspIleAsnSerSerMetThrAsnSerThrAlaAlaSerArgProValThrLeu 100

Qy 901 AAAATCTTGGCACACAATGCTTGGTTGAAGAGACTGATTGGAAGAAGGCAGAAATTTG 960
            ||| :::|||
Db 101 ArgLeuValValProAlaThrGlnCysGlySerLeuIleGlyLysGlyGlyCysLysIle 120

Qy 961 AAGAAAATTGAACATGAAACAGGACCAAGATAACAATCCATCTTGCAGGATTTGAGC 1020
            |||||| |||||| ::: |||
Db 121 LysGluLeuArgGluThrGlyAlaGlnIleLysValAlaGly-----AspMetLeu 138

Qy 1021 ATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTCTGCCAGTGCT 1080
            ::: ||| :::
Db 139 ProAsnSerThrGluArgAlaIleThrIleAlaGlyValProGlnSerValThrGluCys 158

Qy 1081 GAGATAGAGATT--------ATGAAGAAGTGCGTGAGGCCTTTGAAAATGATATG 1128
            |||||||
Db 159 ValLysGlnIleCysLeuValMetLeuGluThrSerGlnSerProGlnGlyArgVal 178

Qy 1129 CTGGCTGTTAACCACAAGGCCAATCTGATCCCCAGGGTTGAACCTCAGCGCACTTGGCATC 1188
            ::: |||
Db 179 MetThrIleProTyrGlnPro--------MetProAlaSerSerProValIle 193

Qy 1189 TTTTCAACAGGACTGTCCGTGCTATCTCCACCAGCGGCCCCCGGGAGCTCCCCCGCT 1248
            |||||| |||
Db 194 CysAlaGlyLyGlnAspArgCysSerSerAspAlaAlaGly------ 206

Qy 1249 GCCCCCTACCACCCCTTCACTCACCCAC---------TCCGGATACTTC 1287
            ::: |||||||
Db 207 --------TyrProHisAlaThrHisAspLeuGluGlyProProLeuAspAlaTyr-- 222

Qy 1288 TCCAGCTGTACCCCCATCACCAGTTTGGCCCG----------------- 1320
            :::   :::
Db 223 --SerIleGlnIleGlyHisHisSerProLeuAspLeuAlaLysLeuAsnGlnVal 241

Qy 1321 ------------------TTCCCGCATCATCAC----- 1335
                              |||||||
Db 242 AlaArgGlnSerHisPheAlaMetMetHisGlyGlyThrGlyPheAlaGlyIleAsp 261

Qy 1336 --TCCTATCCAGAG----------------CAGGAGATT 1356
              |||||| |||||||
Db 262 SerSerSerProGluValLysGlyTyrTrpTrpAlaSerLeuAspAlaSerThrGlnThrThr 281

Qy 1357 GTGAATCTCTTCATCCCAACCCAGCGTGTGGGGCCATCATCGGGAAGAGGGGCCACAC 1416
            :::||||||| :::||||||
Db 282 HisGluLeuThrLeuProAsnAsnLeuIleGlyCysIleIleGluGlyArgGlnGlyAlaAsn 301

Qy 1417 ATCAAACAGTTGGCAGATTCGCCGGAGATTCGCCCTCCATCAAGATTGCC---CCTGCCGGAAGGC 1473
            |||||| :::   :::|||||||
Db 302 IleAsnGluIleArgGlnMetSerGlyAlaGlnIleIleLysAlaAsnProValGluGlyGly 321

Qy 1474 CCAGAGACGCTCAGCGAAAGAGATGGTGCATCATCACCGGGCCACCGGAGCCCAGTTCAGGCC 1533
            :::   :::||||||
Db 322 Ser------SerGlyArgGlnValThrIleThrGlyValThrIleThrGlySerAlaAlaSerIleSerLeuAla 339

Qy 1534 CAGGGACGGTCTTTGGGAAACTGAAAGAGGAA 1566

```
Db  768 --------------------------------------SerSerValArgAspAlaLeuIleGlnIle-ValLeuAr 780
             |||||                                          ::||
Qy  1825 GAGCAGAAATACCCTCAGGAGTGCGCCTCACAGCGCAGCAAGCGAGGCTCCACAGCCAC 1884
             |||                                                :::
Db  780 gLeuArgGluAspValLeuGlyAspLysAspSerValAlaThrArgLysProProAla-- 799
             ||||||
Qy  1885 CAGGCAAAACAACGGATGAATGTAGC 1909
             |||||||
Db  800 ------ArgThrAspAsnCysSer 805

RESULT 12
S46109
hnRNP complex protein homolog YBR233w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YBR1531
C:Species: Saccharomyces cerevisiae
C:Date: 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 19-Apr-2002
C:Accession: S46109
R;Dubois, E.; El Bakkoury, M.; Glansdorff, N.; Messenguy, F.; Pierard, A.; Scherens, B.;
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45782
A:Accession: S46109
A:Molecule type: DNA
A:Residues: 1-413 <DUB>
A:Cross-references: EMBL:Z36101; NID:g536631; PIDN:CAA85196.1; PID:g536633; MIPS:YBR233w
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:PBP2
A:Cross-references: SGD:S0000437
A:Map position: 2R
C:Keywords: nucleus

Alignment Scores:
Pred. No.:                    3.73e-08        Length:          413
Score:                        235.00          Matches:         102
Percent Similarity:           40.20%          Conservative:    56
Best Local Similarity:        25.95%          Mismatches:      151
Query Match:                  3.84%           Indels:          84
DB:                           2               Gaps:            16

US-09-270-437D-6 (1-3412) x S46109 (1-413)

Qy  523 ATTTCCTACATCCGGATGAAGAGGTGAGCTCCCCTGCCCCCTCAGCGAGCGAGGCGT 582
            ||:::|  |||:|||                            :::||
Db  10 IleThrThrProThrThrValGlyThrLeuValSerProAsnThrLeuLysArg--LysLys 28
Qy  583 GGGGGACCACTCTTCCCGGGAGCAA----- 606
            |||
Db  29 GlyGluAspThrSerGluGluGlnLeuGluAlaGluIleLysArgValAlaLeuLysAspAsp 48
Qy  607 --GGCCACGCCCCTGGGGCCACTTCCTCAGGCCGGCAGCAGATT-----GATTCCCG 654
            |||
Db  49 AlaAspSerHisSerAspAsnAspHisAspSerProAspAsnValProSerAspValHis 68
Qy  655 CTGCGGATCCTGGTCCCACCCAGTTGTTGTGGCCATCATCGGAAAGGAGGGCCTTGACC 714
            ||  :::                             :::|||
Db  69 LeuArgMetLeuCysLeuLeuValLysHisAlaSerLeuIleValGlyLysHisLysGlyAlaThr 88
Qy  715 ATAAAGAAACATCACTAAGCACCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCT 774
            |  ::  :::
Db  89 IleSerArgIleLysSerGluThrSerAlaArgIleAsnIleSerArgAsnIleArg-- 107
Qy  775 GGAGCTGCAGAGAAGAAGGCTGTCACCATCCACCCCAGAGAGGGACTTCTGAAGCATGC 834
            ||                                           :::
Db  108 GlyValProGluArgIleValTyrValArgGlyThrCysAspAspValAlaLysAlaTyr 127
Qy  835 CCGCATGATTCTTGAAATCATCGAGAAAGAGAGGCAGCAGATGAGACCCAAACTAGCCGAA------ 888
            ||||                                                   :::
Db  128 GlyMetIleValArgAlaLeuGlnLeuGluGluHisGlyAsnGlyAsnGlyAspIle 147
Qy  889 GAGAGTTCCTGGAAATCTTGGCACCACCAATGGCTTGGTTGGAAGACTGGAAAAGAA 948
            |||||                                    :::|||
Db  148 GluIleSerIleAsnLeuLeuIleProHisHisLeuMetGlyCysIleIleIleGlyLysArg 167
```

```
Db  422 -----------------------TyrAspAlaIleValAlaAlaSerAspGluGluAsnAsnThrLysThr 439
            |||||                                        ::||||
Qy  928 ------------------------------------------------GGA 930
Db  440 AsnValAspArgAspAspAsnLysGluCysArgLeuLeuValProPheSerGlnSerSer 459
            |||||||
Qy  931 AGACTGATTGGAAAAGAAGGCAGAAATTTGAAGAACATGAAACAGGGACCAAG 990
            |||:::  |||  |||::||
Db  460 SerLeuIleGlyLysAlaGlyIleGluAsnIleLysArgIleArgArgArgAlaSer 479
Qy  991 ATAACAATCTCATCTTTGCCAGGATTTGAGCATATACAACCCGGAAAGAACCATCACTGTG 1050
            |||  ::   |||||                        :::||
Db  480 ValLysValValSer-------LysAspValSer------AspProSerHisValCysAlaMet 496
Qy  1051 AAGGGCACAGTTGAGCCTGTGCCAGTGCTGAGATAGAGATTATGAAGAGCTCGTGAG 1110
            |||                                     |||::
Db  497 GluTyrAspAsnValValIleSerGlyValGlyGluProGluSerValLysGlnAlaLeuPhe 516
Qy  1111 GCCTTTGAAAATGATATGCTGGCTGTTAACCAACAAGCCAATCTG-------------- 1155
            |||                                      ::|||
Db  517 AlaValSerAlaIleAlaMetTyrLysIleAsnProArgGluAsnIleProLeuAspSerThr 536
Qy  1156 -------ATCCCAGGGGTTGAACCTC-------------------- 1173
            :::|||                          :::||
Db  537 SerGlnAsnAspValProAlaAlaSerValIleValProSerAspLeuSerAsnSerValTyr 556
Qy  1174 AGCGCACTTGGCATCTTTCAACAGGACTGTCCGTGCTATCTCCACCAGCAGGGCCCCGC 1233
            |||                                         ::|||
Db  557 ProGlnThrGlyPheTyrSerAsnGlnAspHisIleLeuGlnGlnGlyValAlaGlyValPro 576
Qy  1234 GGAGCTCCCCCGCTGCCCCCTACACCCCTTCACTACCCACTCCGGA--------------- 1281
            |||                                       :||
Db  577 SerTyrPheAsnAlaLeuSerValSerAspPheGlnGlyTyrAlaGluThrAlaAlaAsn 596
Qy  1282 --------TACTTCTCCAGGCTGTACCCCCATCACCAGTTGGCCCGTTCCGCAT-------- 1329
            |||                                         ||||||
Db  597 ProValProValPheAlaSerSerLeuProValThrHisGlyPheGly-------------- 612
Qy  1330 CATCACTCTTATCCAGAGCAGGAGATTGTGAATCTTCTCATCCCAACCCAGCCTGTGGGC 1389
            |||                                        :::
Db  613 GlySerSerArgSerGluGluValPheLysValLeuCysProLeuCysAsnIleMet 632
Qy  1390 GCCATCATCCGGAAGAAGGGGCCACATCAAACAGTCTGGCGAGATTCGGCCGGAGCCTCT 1449
            |||:::||                                       ||||
Db  633 ArgValIleGlyLysLysGlyIleGlySerThrIleArgIleLysArgGluArgGlySerCys 652
Qy  1450 ATCAAGATTGCCCCTGCGGAAGCCCCAGACGTCAGCGAAAGGATGGTCATCATCCACCGGG 1509
            |||                                           :::
Db  653 IleGluValAsnAspSerArgThrLysCysGlyAspAspValCysValIleIleIleValThr 672
Qy  1510 CCACCCGGAGCCCAGTTCAAGCCGTCAAGGCCGCAGGGACGGATCTTT-------------- 1548
            |||                                            :::
Db  673 AlaThrGluIleLeuPheCysCysLeuSerThrProPheMetGlnSerProAsp 692
Qy  1549 --------GGGAAACTGAAAGAGAGGAAAACTTCTTTAACCCCCAAA 1584
            |||                           :::
Db  693 AspMetLysSerMetAlaValGlyAsnValLysMetGlnLeuLeuLeuLeuLeuAsnGluTyrIleLeuAsnAspGlu 712
Qy  1585 GAAGAAGTGAAGCTGGAAGCCGCCATATCAGAGTGCCCCTCTTCCACAGCTGGCCGGGTGATT 1644
            |||                                         :::
Db  713 IleGluValAlaAsnAspSerArgThrLysCysGlyAspAspValCysValIleIleIleValThr 732
Qy  1645 GGCAAAGGTGGCAAGACCGTGAACCGTTAACCAGTGCAGAAGTCATCGTG 1704
            |||||                                       :::
Db  733 GlyLysSerGlyValGlySerValIleAsnAlaAsnIleCysValIleIle 752
Qy  1705 CCTCGTGACCAAACGCCAGACGTCAGCGAAATGATGAAAATATCGGGCACTTC 1764
            |||                                    :::|||
Db  753 SerLysGlyLysLysAspAsp--------LeuValGlyValSerGlyGluVal 767
Qy  1765 TTTGCTAGCCAGACTGCACAGCCGCAAGATCAGGAAATTGTACAACAGGTGAAGCAGCAG 1824
            :::|||                                         :::
Db  768 GluIleSerIleAsnLeuLeuIleProHisHisLeuMetGlyCysIleIleIleGlyLysArg 780
```

```
Alignment Scores:
Pred. No.:              1.62e-07                    Length:        479
Score:                  226.00                      Matches:        97
Percent Similarity:     39.18%                      Conservative:   95
Best Local Similarity:  19.80%                      Mismatches:    148
Query Match:            3.69%                        Indels:        150
DB:                     2                           Gaps:           18

US-09-270-437D-6 (1-3412) x C86275 (1-479)

Qy    583  GGGGACCACTCTTCCCGGGAGCAAGGCCACGCCCTGGGGCACTTCCAGGCCAGACAG  642
                |||||| |||     :::   |||             ||| |||||||||  |||
Db     10  ----------GlyLysArgSerAsnLeuGlnSerGluPheThrGlyAsnGlyLySerLysArgArgAsn   29

Qy    643  ATT----------------------------------------------------GATTTCCCGTCGCGATCCTG  666
           |||                                                     ||| |||
Db     30  LeuHisAspGluThrAspGlnAsnValIleAlaAlaSerGluAspThrValTyrArgTyrLeu   49

Qy    667  GTCCCCACCCAGTTGTTGTGTGCCATCATCGGAAAGAGGAGGCTTGACCATAAAGACAATC  726
                ||| ||| ::  |||::|||:::|||||||||||    ||||
Db     50  CysProValLysThrGlySerIleIleGlyLysGlyIleGlyIleAlaAlaLysGlnIle   69

Qy    727  ACTAAGCAGACCAGTCCCGGGTAGATATCCATAGAAAGAGAACTCTGGAGCTGCAGAG  786
                :::|||||||||||||:::||||||||          :::        :::
Db     70  ArgSerGluThrLysSerAsnMetArgIleAsn---GluAlaLeuProGlyCysGluGlu   88

Qy    787  AAGCCTGTCACCATCCATCCCACCCACCCCCAGAGGGACTTCT----------------  825
                :::::|||||:::::|||    :::|||||
Db     89  ArgValValThrMetTyrSerThrAsnGluGluLeuAsnHisPheGlyAspAspGlyGlu  108

Qy    826  ----------------------------------------GAAGCATGCCCCATGATTCTTGAAATCATGCAGAAAGAGGCA  867
                                                    ||| :::|||
Db    109  LeuValCysProAlaLeuAspAlaLeuPheLysValHisAspMetValValAlaAspAla  128

Qy    868  ------------------------GATGAGACCAAACTAGCCGAAGAG----ATTCCTCTG  900
                                    ||| ||||        |||
Db    129  AspGlnAspAspGlyThrAspAspAsnAspLeuGlyLysGlyGlnThrValThrVal  148

Qy    901  AAAATCTTGGCACCACCAAGTCGTTGGTTGGAAGACTGATTGGAAAAGAGGCAGAAATTG  960
                :::||:||        :::   |||||||||
Db    149  ArgMetLeuValProSerAspGlnIleGlyCysValIleGlyLysGlyGlnValIle  168

Qy    961  AAGAAAAATTGAACATGAAACAGGGACCAAGAAACAATCTCATCTTTGCAGGATTTGAGC  1020
                :::      :::||||      :::
Db    169  GlnAsnLeuArgAsnAspThrAsnAlaGlnIleArgVal---------------------  181

Qy   1021  ATATACAACCCGGAAAGAGACCATCACTGTGAAGGGCCTGTGCC--------------  1074
                                                |||||||||
Db    182  --------------------------------IleLysAspHisLeuProAlaCysAlaLeuThr  192

Qy   1075  ------------------------------------------------AGTGCTGAG  1083
                                                            ||| |||
Db    193  LeuSerHisAspGluLeuLeuLeuLeuIleLeuGlyGluProLeuValValArgGluAlaLeu  212

Qy   1084  ATAGAGATTATGAGAAGCTGCCTGAG----------GCCTTTGAAAATGATATGCTGGCT  1134
                                            ||||||
Db    213  TyrGlnValIleAlaSerLeuLeuHisAspAsnProSerArgPheGlnHisIleLeuLeuSerLeuSer  232

Qy   1135  GTTAACCAACAAGCCAATCTGATCCCAGGGTTGAACCTGCCACTTGGCATCTTTTCA  1194
                :::                     :::|||||||||
Db    233  SerSerSerSerMetHisGlnProGlyAlaMetLeuMetSerAlaAlaLeuThrSer  252

Qy   1195  ACAGGACTGTCCGTCATCTCCACCAGCAGGCCCCCGCTGCCCC  1254
                ::|
Db    253  Ser--------------------------------------  253

Qy   1255  TACCACCCTTCACTACCCACTCCGGATACTTCTTCCAGCACTACCCCCATCACCAGTT  1314
                                                ||| :::
Db    253  -----------------------------------------  253

Qy   1315  GGCCCGTTCCCGCATCATCACTCTTAT------------CCAGAGCAGGAG  1353
                                          ::: |||
Db    254  -----------------HisArgAsnTyrAlaValArgArgAspIleAlaAspIleAlaArgGlu  268
```

```
Qy    949  GGCAGAAATTTGAAGAAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTG  1008
                |||||| :::||||||||           ||| :::    :::
Db    168  GlySerArgLeuArgGluIleGluAspLeuAspLeuSerAlaAlaAlaLysPheAlaSerProAsn  187

Qy   1009  CAGGATTTGAGCATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCC  1068
                |||           :::|||   ::::::|||
Db    188  Gln------LeuLeuLeuSerAsnAspAsnArgIleLeuThrLeuAsnGlyValProAspAla  205

Qy   1069  TGTGCCAGTGCTGAGATAGAGATTATGAGAAGCTG----------------------  1104
           |||
Db    206  IleHisIleAlaThrPheTyrIleSerGlnThrLeuLeuLeuAsnPheGlnMetGluSerPro  225

Qy   1105  --------------CGTGAGGCCTTTGAA----------------------AATGATATGCTGGCT  1134
                          :::                             |||
Db    226  GlnLysAsnValLysArgSerIleTyrTyrGlnProThrGlnProGlnPheAsnSerValIleIle  245

Qy   1135  GTTAACCAACAAGCCAATCTGATC----------------------------------  1158
           :::                     |||
Db    246  AspHisSerGlnProAsnThrIlePhePheHisGlnIleArgAsnHisGlnTyrHisProSerAsp  265

Qy   1159  --------CCAGGGGTTGAACCTGCCACTTGGCCATCTTTTCAACAGGA  1200
                   |||||||||               |||
Db    266  LysLeuLeuSerTyrLysProAsnLysAsnLeuPro------------IleSerSerThrLeu  282

Qy   1201  CTGTCCGTGCTATCTCCACCAGCAGGCCCCCGCTGCCCCTACCAC  1260
                |||||                         :::
Db    283  LeuSerMetAlaThrProGlnTyrThrThrAlaSerValAlaAsnAlaThrAlaPheGln  302

Qy   1261  CCC----------------TTCACTACCCACTCCGGATACTTCTTCCAGCCTGTACCCC----------  1302
           |||                |||
Db    303  ProAsnPheValIleProAsnValThrValLeuAspGlyProValIleSer---------  319

Qy   1303  CATCACCAGTTGGCCCGTTCCCGCATCATCAC----------TCTTATCCAGAGCAGGAG  1353
           |||                             :::|||||
Db    320  --------------ProAlaProGlyAsnHisLeuLeuMetAsnPheValGlnGlnGlu  334

Qy   1354  ATTGTGAATCTTCATCCAACCCAGGCTGTGGGCGCCATCATCGGGAAGAAGGGGGCA  1413
                                              ||||||||||||
Db    335  Ile--------PheIleAspGluLysPheValGlyAsnValIleIleGlyLysGlyAspGlyLys  351

Qy   1414  CACATCAAACAGTCCGAGATTCGCCGAGACGATTGCC---CCTGCCGAA  1470
           ||||||                            :::
Db    352  HisIleAsnSerValLysSerGluValThrGlyCysSerIleIleIleGluGlnProValGlu  371

Qy   1471  GGGCCAGACGTCAGCGAAAGAGATGGTCATCATCACCGGG  1509
                |||||||
Db    372  GlySer------SerGluArgArgLeuThrIleArgGly  382
```

```
RESULT 13
C86275

hypothetical protein (imported) - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C;Accession: C86275
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Cross-references: GB:AE005172; NID:g5080792; PIDN:AAD39302.1; GSPDB:GN00141
A;Accession: C86275
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-479 <STO>
C;Genetics:
A;Map position: 1
```

Qy   682 GTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAGACACATCTAGGAGACACCCAG 741
                 |||||| ::: |||||||:::::|||||||||||
Db   61 AlaGlySerIleIleGlyLysGlyLysGlyLysGlnThrIleValGlnLeuGlnLeuGluThrGlyLysGly 80

Qy   742 TCCCGGGTAGATATCCATAGAAAAGAGAACTCT-------------------GGAGCTGCAGAG 786
                  ::: ||||||||||||::: |||||
Db   81 AlaThrIleLysLeuSerLysLysSerLysAspPheTyrProGlyThrThrGlu 100

Qy   787 AAGCCTGTCACCATCGCCATCATCGCCGGGGACTTCTGAAGCATGCCGCCATGATTCTT 846
                 ||||||||::: |||||:::||||||
Db   101 ArgValCysLeuIleGlnGlyThrValGluAlaLeuAsnAlaValHisGlyPheIleAla 120

Qy   847 GAA-------------------------------ATCATGCAG 858
                 |||
Db   121 GluLysIleArgGluMetProGlnAsnValAlaAlaLysThrGluProValSerIleLeuGln 140

Qy   859 AAAGAG-----GCAGTGAGACCAAACTAGCCGAAGAGATTCCT------------------- 897
                 :::   ::: ||| ::::|||
Db   141 ProGlnThrThrValAsnProAspArgIleLysGlnThrLeuProSerSerProThrThr 160

Qy   898 --------------------------------------CTGAAAATCTTG 909
                                                       :::||||||
Db   161 ThrLysSerProSerAspProMetThrThrSerArgAlaAsnGlnValLysIleIle 180

Qy   910 GCACAACAATGGCTTGGTTGGTGGAAGAAGGACTGATTGGAAAAGAAGGCAGAAATTTGAAGAAATT 969
                 |||||| :::||:::|||||||::: :::|||
Db   181 ValProAsnSerThrAlaGlyLeuIleLeuGlyLysGlyLysGlyAlaThrValLysAlaVal 200

Qy   970 GAACATGAAACAGGGACCAAGATAACAATCCATCTTCAGGATTTGAGCATATACAAC 1029
                 :::||| ::: |||:::||
Db   201 MetGluGlnSerGlyAlaTrpValGlnLeuSerGlnLysProAsp---GlyIleAsnLeu 219

Qy   1030 CCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGCCTGCCCAGTGCAGATGAGAG 1089
                 ||| :::||:::||||||
Db   220 GlnGluArgValValThrValSerGlyValGluProGluGlnAsnArgLysAlaValGluLeu 239

Qy   1090 ATTATGAAGAAGCTGCCTGAG-------------------------------------- 1110
                 ||| ::::|||||::::||||
Db   240 IleGluIleGlnLysIleGlnLysGluAspProGlnSerGlySerCysLeuAsnIleSerSerTyrAla 259

Qy   1110 --- 1110

Db   260 AsnValThrGlyProValAlaAsnSerAsnProThrGlySerProTyrAlaAsnThrAla 279

Qy   1110 --- 1110

Db   280 GluValLeuProThrAlaAlaAlaAlaAlaGlyLeuLeuGlyHisAlaAsnLeuAlaGly 299

Qy   1111 ---------GCCTTTGAA---------AATGATATGCTGGCTGTTAAC 1140
                        ||||| ||::::|||:::
Db   300 ValAlaAlaPheProAlaValLeuSerGlyPheThrGlyAsnAspLeuValAlaIleThr 319

Qy   1141 CAACAGCCAATCTGATCCCA------GGGTGAACCTCAGCGCACTTGGCATCTTTTCA 1194
                 |||||| ::: |||||
Db   320 SerAlaLeuAsnThrLeuAlaSerTyrGlyTyrAsnLeuAsnThrLeuGlyLeu------------------- 337

Qy   1195 ACAGGACTGTCCGTGCTATCTTCCACCAGCAGCCCCGCGAGCT------------------- 1239
                 ||||||| :::|| ||||||
Db   338 ---GlyLeuSer-----------GlnAlaLaAlaAlaThrGlyAlaLeuAlaLeuAlaAlaAlaAla 352

Qy   1240 --------------CCCCGGTGCCCCTACCACCCCTCACTACCCAC------------------- 1275
                              ||| |||||
Db   353 AlaSerAlaAsnProAlaAlaAlaAlaAsnLeuLeuAlaThrTyrAlaSerGluAla 372

Qy   1275 --- 1275

Db   373 SerAlaSerGlySerThrAlaGlyThrAlaGlyThrPheAlaLeuGlySerLeuAla 392

Qy   1276 ---------------------------TCCGGATACTTCTTCCAGGCTGTACCCCCATCACCAGTTTGGC 1317
                              :::||||
Db   393 AlaAlaThrAlaAlaThrAlaAsnGlyTyrPheGlyAlaAlaSerProLeuAlaAlaSerAla 412

Qy   1318 CCGTTCCCGCCATCATCACTCTTATCCA--GAGCAGGAGATTGTGAATCTTCATCCCA 1374
                 |||||| ||| :::
Db   413 ...

Qy   1354 ATTGTGGAATCTTCATC---CCAACCCAGGCTGTGGGCGCCATCATCGGGAGAAAGGGG 1410
                 ||||| ::: ||| :::|||||||
Db   269 PheCysValCysPheIleCysProAlaGlyGluAsnValGlyValIleGlyValIleGlyValIleGlyVallGly 288

Qy   1411 GCACACATCAAACAGCTGGCGAGATTCGCCGGAGCCTCTCATCCAAGATTCCCCTGCGGAA 1470
                 ||||| ::: |||||||:::|||
Db   289 GlyPheIleAsnGlnIleLeuArgGlnGluThrGlyIleValAlaThrIleArgValAlaAsnThrSerGlu 308

Qy   1471 GGCCCAGAGGTCAGCGAAAGGATGGTCATCATCACCGGGGCCACCGGAA------------------- 1518
                 ||| ::: |||
Db   309 ThrAspAsp-----AspAspGlyCysIleIleGluIleSerGluPheSerSerGluLysLysGluGlyTyrGluTyrAsp 326

Qy   1519 ---------GCCCAGTCAAGGCCCCAGGACGGATCTTTGGGAAACTG-- 1557
                        ||| ||| |||
Db   327 GlnSerProAlaValAlaAsnAlaAlaIleAlaArgLeuGlnLeuArgCysSerSerValGlyGly 346

Qy   1558 AAAGAGGAAAACTTCTTAACCCCAAAGAAGAGTGAAGCTGGAAGGCGCATATCAGAGTG 1617
                 ||||| ::: ||:::
Db   347 --------------AspGluLaAlaIleSerGluThrArgLeuLeuLeuVal 360

Qy   1618 CCCTTCCACAGCTGCCGGGTGATTGGCAAAGGTGGCAAGACCGTGAAGACCGTGAAACTGCAG 1677
                 |||||| :::||||||||||:::
Db   361 SerSerSerGlnIleLeuGlyLeuGlyGlyLysGlyLysGlyAlaValIleIleSerGluGluMetArg 380

Qy   1678 AACTTAACCAGTGCAGAAGTC--ATCGTGCCTCGTGACCAAAGCCCAGAT------GAA 1728
                 ::::||| ||| ||::: ||| :::::::|||
Db   381 SerValThrArgAlaAlaAsnIleArgIleLeuArgIleLeuGluIleLysGluIleAsnAspValProLysIleAlaArg 400

Qy   1729 AATGAGGAAGTGATCGTCAGAAAATATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGC 1788
                 ||| ||::: :::|||
Db   401 GluAspGluGluValGlnIleValGlnIleLeuThrGly------SerProAspAlaAlaMetLys 417

Qy   1789 AAGATCAGGGAAATTGTACAACAGTGAAG 1818
                 :::::::||||||| ::::::::
Db   418 AlaLeuThrGlnIleValIleLeuLeuArgLeuArg 427

Alignment Scores:
Pred. No.:   US-09-270-437D-6 (1-3412) x I38489 (1-510)
Score: 1.65e-07
Percent Similarity: 226.00
Best Local Similarity: 34.29%
Query Match: 21.56%
DB: 3.69%

Length: 510
Matches: 105
Conservative: 62
Mismatches: 158
Indels: 162
Gaps: 16

Qy   562 CCCCCTCAGCGAGCCCAGCGTGGGAGACCACTCTTCCCGGAGCAAGGCCACGCCCCTGGG 621
                 ||||| ::: |||
Db   22 ProProAspSerArgLysArgProLeuGluAlaProGluAlaGly---SerThrLys 40

Qy   622 GGCACTTCTCAGGCCAGACAGATTGATTCCCCGCTGCGGATCCTGGGTCCCCACCCAGTT 681
                 |||::::   ::: |||
Db   41 ArgThrAsnThrGlyValGlyGluAspGlyGlnTyrPheLeuLysValLeuIleLeuLysValLeuIleLeuProSerTyrAla 60

```
Db   413  IleLeuGlyThrGluLysSerThrAspGlySerLysAspValValGluIleAlaValPro 432
               |||          ::::::::|||     :::|||
Qy  1375  ACCCAGGCTGTGGGGCCATCATCGGGAAGAAGAGGGGCCACACAAACAGCTGCCGAGA 1434
              |||||:|||||||||::|||||||||        :::|||
Db   433  GluAsnLeuValGlyAlaIleLeuGlyLysGlyThrLeuValGluTyrGlnGlu 452
              |||||:|||||||||::|||:     :::|||
Qy  1435  TTCGCCGGAGCCTCATCAAGATTGCCCCTGCCGAAGGC--------CCAGACGTCAGC 1485
              ||||||:|||||:::        |||
Db   453  LeuThrGlyAlaArgIleGlnIleSer---LysLysGlyGlyPheValProGlyThrArg 471
              |||||||:|||||||||         |||
Qy  1486  GAAAGGATGGTCATCATCACCGGGCACCGGGAAGCCAGTTCAAGGCCCAGGGACGGATC 1545
              ||||||:|||     :::|||
Db   472  AsnArgLysValThrIleThrGlyThrProAlaIleAlaThrGlnAlaAlaGlnTyrLeuIle 491
              |||||:|||          :::|||
Qy  1546  TTTGGGAAACTGAAAGAGGAA 1566
              ::::::
Db   492  ThrGlnArgIleThrTyrGlu 498

RESULT 15
S42471
hnRNP protein E2 - human
N;Alternate names: nucleic acid-binding protein; protein PCBP-2
C;Species: Homo sapiens (man)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 05-Nov-1999
C;Accession: S65679, S42471
R;Leffers, H.; Dejgaard, K.; Celis, J.E.
Eur.J.Biochem. 230, 447-453, 1995
A;Title: Characterisation of two major cellular poly(rC)-binding human proteins, each co
A;Reference number: S65678; MUID:95331278; PMID:7607214
A;Accession: S65679
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-365  <LEF>
A;Cross-references: EMBL:X78136; NID:g460772; PIDN:CAA55015.1; PID:g460773
A;Experimental source: AMA cells (transformed human amnion cells)
A;Note: submitted to the EMBL Data Library, March 1994
C;Keywords: RNA binding

Alignment Scores:
Pred. No.:         1.63e-07        Length:           365
Score:             225.50          Matches:          89
Percent Similarity: 41.40%         Conservative:     65
Best Local Similarity: 23.92%      Mismatches:       153
Query Match:        3.68%          Indels:           65
DB:                 2              Gaps:             14

US-09-270-437D-6 (1-3412) x S42471 (1-365)

Qy   601  GAGCAAGGCCACGCCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGG 660
              ::: |||||:::||||||||          ::              :::|||
Db     2  AspThrGlyValIleGluGlyGlyLeuAsn---------ValThrLeuThrIleArg 17
              |||::|||        ::               :::|||
Qy   661  ATCCTGGTCCCCACCCAGTTTGTGTGGTGCCATCATCGGAAAGAGGAGGCTTGACCATAAAG 720
              :::||:|||::        ::: |||||||||::::|||
Db    18  LeuLeuMetHisGlyLLysGluValGlySerIleIleGlyLysLysGlyGluSerValLys 37
              |||:||||:        :::::|||
Qy   721  AACATCACTAAGCACAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAACTCTGGAGCT 780
              ::: ||||::        |||
Db    38  LysMetArgGluGluSerGlyAlaArgIleAsnIle--------SerGluGlyAsnCys 54
              |||::|::||        :::|||
Qy   781  GCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGACTTCTGAAGCCATGCCCATG 840
              |||:::||:||        ::|||
Db    55  ProGluArgIleIleThrLeuAlaGlyProThrAsnAlaIlePheLysProAlaMet 74
              |||:|||:|||        :::|||
Qy   841  ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTA-------------.882
              ::: :::::::::::        ::: |||
Db    75  IleIleGluAspLeuGluGluAspGluSerSerMetThrAsnSerThrAlaAlaSer 94
              |||:::|||        ::|||
Qy   883  GCCGAAGAGATTCCTGAAAATCTTGGCACACAAATGGCTTGGTTGGAAGACTGATTGGA 942
              |||:|:|||        :::|||
Db    95  ArgProValThrLeuArgLeuValValProAlaSerGlnCysGlySerLeuIleGly 114
```

```
                        GenCore version 5.1.6
                Copyright (c) 1993 - 2004  Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:       July 16, 2004, 10:30:58 ; Search time 31.5 Seconds
                                    (without alignments)
                              11280.214 Million cell updates/sec

Title:          US-09-270-437D-6
Perfect score:  6121
Sequence:       1 ggcagcgaggaggcgagga........aacctgaaatgttattt 3412

Scoring table:  BLOSUM62
                Xgapop  10.0 ,  Xgapext   0.5
                Ygapop  10.0 ,  Ygapext   0.5
                Fgapop   6.0 ,  Fgapext   7.0
                Delop    6.0 ,  Delext    7.0

Searched:       141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:    283362

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09270437/runat_16072004_113049_13315/app_query.fasta_1.3591
-Db=SwissProt_42 -QFMT=fastan -SUFFIX=rsp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09270437 @CGN_1_1_46 @runat_16072004_113049_13315 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      SwissProt_42:*

    Pred. No. is the number of results predicted by chance to have a
    score greater than or equal to the score of the result being printed,
    and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Match	Query Match Length	DB	ID	Description	
1	278	4.5	721	1	FUB2_RAT	Q99pf5	rattus norv
2	262.5	4.3	572	1	FUB3_HUMAN	Q96l24	homo sapien
3	260.5	4.3	769	1	ZBP2_CHICK	Q8uvd9	gallus gall
4	251	4.1	707	1	FUB2_HUMAN	Q92945	homo sapien
5	245.5	4.1	339	1	PCB3_HUMAN	P57721	homo sapien
6	244.5	4.0	339	1	PCB3_MOUSE	P57722	mus musculu
7	244	4.0	651	1	FUB1_MOUSE	Q91wj8	mus musculu
8	242.5	4.0	643	1	FUB1_HUMAN	Q96ae4	homo sapien
9	238.5	3.9	356	1	PCB1_RABIT	O19048	oryctolagus
10	238.5	3.9	606	1	Y475_ARATH	P58223	arabidopsis
11	238	3.9	492	1	NOA2_HUMAN	Q9unw9	homo sapien
12	235.5	3.8	474	1	NOA1_RAT	Q80wa4	rattus norv
13	235	3.8	413	1	PBP2_YEAST	P38151	saccharomyc
14	234.5	3.8	356	1	PCB1_HUMAN	Q15365	homo sapien
15	231.5	3.8	403	1	PCB4_HUMAN	P57723	homo sapien
16	231.5	3.8	403	1	PCB4_MOUSE	P57724	mus musculu
17	226.5	3.7	493	1	NOA1_MOUSE	Q9jkn6	mus musculu
18	226	3.7	510	1	NOA1_HUMAN	P51513	homo sapien
19	225.5	3.7	365	1	PCB2_HUMAN	Q15366	homo sapien
20	225	3.7	362	1	PCB2_MOUSE	Q61990	mus musculu
21	203.5	3.3	1268	1	VGLN_HUMAN	Q00341	homo sapien
22	201.5	3.3	1270	1	VGLN_CHICK	P81021	gallus gall
23	187	3.1	606	1	TDRH_HUMAN	Q9y2w6	homo sapien
24	185.5	3.0	463	1	ROK_HUMAN	Q07244	homo sapien
25	185.5	3.0	463	1	ROK_RABIT	O19049	oryctolagus
26	184.5	3.0	629	1	PAB2_ARATH	P42731	arabidopsis
27	179	2.9	5262	1	MLL2_HUMAN	O14686	homo sapien
28	176.5	2.9	470	1	NR54_HUMAN	Q15233	homo sapien
29	176	2.9	1222	1	S160_YEAST	P06105	saccharomyc
30	174.5	2.9	817	1	VRP1_YEAST	P37370	saccharomyc
31	174.5	2.9	1781	1	AK12_HUMAN	Q02952	homo sapien
32	169	2.8	1943	1	PC15_MOUSE	Q99pj1	mus musculu
33	168.5	2.8	5147	1	PCLO_HUMAN	Q9y6v0	homo sapien
34	165	2.7	2161	1	SHK1_HUMAN	Q9y566	homo sapien
35	164	2.7	2167	1	SHK1_RAT	Q9wv48	rattus norv
36	164	2.7	5703	1	MU5B_HUMAN	Q9hc84	homo sapien
37	162	2.6	653	1	PABP_SCHPO	P31209	schizosacch
38	161	2.6	660	1	YHL1_EBV	P03181	epstein-bar
39	158	2.6	2842	1	APC_RAT	P70478	rattus norv
40	157	2.6	1687	1	VIT2_FUNHE	Q98893	fundulus he
41	156	2.5	1329	1	KF10_HUMAN	Q9p218	homo sapien
42	155.5	2.5	2517	1	NCR2_HUMAN	Q9y618	h nuclear r
43	155	2.5	381	1	YBD2_YEAST	P38199	saccharomyc
44	153.5	2.5	636	1	PAB1_HUMAN	P11940	homo sapien
45	153.5	2.5	636	1	PAB1_MOUSE	P29341	mus musculu

ALIGNMENTS

```
RESULT 1
FUB2_RAT
ID   FUB2_RAT    STANDARD;    PRT;    721 AA.
AC   Q99PF5;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Far upstream element binding protein 2 (FUSE binding protein 2) (KH
DE   type splicing regulatory protein) (KSRP) (MAP2 RNA trans-acting
DE   protein 1) (MARTA1).
GN   FUBP2 OR KHSRP.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., SEQUENCE OF 73-87, 89-108 AND 475-486, SUBCELLULAR
RP   LOCATION, AND FUNCTION.
RC   TISSUE=Brain;
RX   MEDLINE=22246918; PubMed=12358751;
RA   Rehbein M., Wege K., Buck F., Schweizer M., Richter D., Kindler S.;
RT   "Molecular characterization of MARTA1, a protein interacting with the
RT   dendritic targeting element of MAP2 mRNAs.";
RL   J. Neurochem. 82:1039-1046(2002).
CC   -!- FUNCTION: Part of a ternary complex that binds to the downstream
CC       control sequence (DCS) of the pre-mRNA. Mediates exon inclusion in
CC       transcripts that are subject to tissue-specific alternative
CC       splicing. May interact with single-stranded DNA from the far-
CC       upstream element (FUSE). May activate gene expression (By
CC       similarity). Binds to the dendritic targeting element and may play
CC       a role in mRNA trafficking.
CC   -!- SUBUNIT: Part of a ternary complex containing FUBP2, PTBP1, PTBP2
CC       and HNRPH1 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear. A small proportion is also found in
CC       the cytoplasm of neuronal cell bodies and dendrites.
CC   -!- SIMILARITY: Contains 4 KH domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@sib-sib.ch).
CC
DR   EMBL; AF308818; AAG59811.1; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 4.
DR   SMART; SM00322; KH; 4.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
KW   Transport; mRNA transport; mRNA processing; mRNA splicing;
KW   Transcription regulation; Trans-acting factor; Nuclear protein;
KW   DNA-binding; RNA-binding; Repeat.
FT   DOMAIN      145    209    KH 1.
FT   DOMAIN      234    300    KH 2.
FT   DOMAIN      323    387    KH 3.
FT   DOMAIN      425    492    KH 4.
FT   REPEAT      572    685    4 X 12 AA IMPERFECT REPEATS.
FT   REPEAT      572    583    1.
FT   REPEAT      618    629    2.
FT   REPEAT      644    655    3.
FT   REPEAT      674    685    4.
FT   DOMAIN        7     68    GLY/PRO-RICH.
FT   DOMAIN       69    498    GLY-RICH.
FT   DOMAIN      499    613    ALA/GLY/PRO-RICH.
SQ   SEQUENCE   721 AA;  74226 MW;  482C7A765C60EE4A CRC64;
```

```
Alignment Scores:
Pred. No.:             4.46e-10      Length:         721
Score:                 278.00        Matches:        178
Percent Similarity:    33.89%        Conservative:    87
Best Local Similarity: 22.76%        Mismatches:     251
Query Match:           4.54%         Indels:         266
DB:                    1             Gaps:            37
```

US-09-270-437D-6 (1-3412) x FUB2_RAT (1-721)

```
Qy      1 GGCAGCGGAGGAGGGAGGAGGCGCCGGGACCGGGCCGGGGAGCCGGGGCTCTCGGGG     60
          ||:||::||||||||          ||| ||| |||  ||
Db     25 GlyAlaGlyGlyGlyProProGlyProProGlyProGlyAlaGlyAspArg-GlyGlyGlyGl   44

Qy     61 AAGAGACGGATGAGTGAACAAGCTTACATCGGAGAAGCTTGAGCCCCGTCACCGCCGAC    120
                                                        ---GlyGlyProGl   51
Db     44 yProGlyGly

Qy    121 GACCTCCGGACAGCTCTTTGGGACAGGAGAGTCGCCCTGCCGGGACAGTCCTGCTGAAG    180
          |||            ||||||
Db     51 yGlyGlyGlyAlaSer-GlyGlyProSerGlnProProGlyGlyGlyGlyProGlyIleA   71

Qy    181 TCCGGCTACGGCCCTCGTGACTACCCCGACTACCCCGGAACTGGGCATCCGCGCACCGAGACC   240
          ||:||||||                        ||||:::::
Db     71 rgLys-AspAlaPheAlaAsp----------------------AlaValGlnArgAlaArgGln   84

Qy    241 CCTCTCGGGTAAAGTG-------                                        255
          ::::|    |||.::
Db     85 IleAlaAlaLysIleGlyGlyAspAlaGlyAlaAlaThrThrValAsnAsnAsnThrProAspPhe  104

Qy    256 --------GAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTA    306
                   :::::   |||         ::
Db    105 GlyPheGlyGlyGlyLnLysArgGlnLeuGluAspGlnLysAspSerLysLysLeu    124

Qy    307 AGGAGCAGGAAAATTCAGATTCGA--------AACATCCCTCCCTCACCTG    348
          |||               :::::: ||
Db    125 AlaSerGlnGlyAspSerIleGlySerGlnLeuGlyProIleHisProProProArgThr   144

Qy    349 ---------CAGTGGGAGGTGTTGGATGGACTTTTGGCTCAATATGGGACAGTG---    393
                  ::::||||||.::
Db    145 SerMetThrGluTyrArgValProAspGlyMetVal-----------GlyLeuIleIle   161

Qy    394 ------------GAGAATGTGGAACAAGTCAACACAGACACAGAAACCGCCGTTGTCAAC    441
          |||            ::::::
Db    162 GlyArgGlyGlyGluGlnIleAsnLysIleGlnAsnGlnAspSer-------------   175

Qy    442 GTCACATATGCAACAAGACAAGAGAAGAAGAAGAAATAGCCATGGAGAAGCTAAGCGGGGCATCAG   501
```

```
Db    176 --------------------------------GlyCysLysValGlnIleSerProAspSerGlyGly--  187
                                          |||:::   :::
Qy    502 TTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGATGAAGAGAGTGAGTCCCCTTCG  561
          |||||||||||||||||
Db    188 --------------------------------LeuProGluArgSerValSerLeuThrGly  197

Qy    562 CCCCCTCAGGCGAGCCCAGCGTGG-----------------------GACCACTCTCCGGAGCAA  606
          |||:::                    ||||:::
Db    198 AlaProGluSerValGlnLysAlaGlnLysMetLeuAspAspSerIleValSerArgArgGlyArg  217

Qy    607 GGCCAGCGCCCTGGG-----------------------------GGCACTCTCAG  633
          |||||                                        |||||
Db    218 GlyGlyProProGlyGlnPheHisAspHisAsnAlaAsnGlyGlyGlnThrValGln  237

Qy    634 GCCAGACAGATTGATTCCCGTGCGGATCCTGGTCCCCACCCAGTTGTTGGTGCCATC  693
          ||||::                                       |||:::
Db    238 --------------------------GluIleMetIleProAlaGlyLysAlaGlyLeuVal  249

Qy    694 ATCGGAAAGAAGGGGCTTGACCATAAAGAAGAACATCACTAAGGACACCAGTCCCGGTAGAT  753
          |||||||||              ||||||
Db    250 IleGlyLysLysGlyGluThrIle-------LysGlnLeuGlnGluArgAlaGly  265

Qy    754 ATCCCATAGA----------------------------AAAGAACTCTGGAGCTCCAGAAGAAGCCTGTC  795
          :::                                    ||||
Db    266 ValLysMetIleLeuLeuIleGlnAspGlySerGlnAsnThrAsn----ValAspLysProLeu  284

Qy    796 ACCATCCATGCCACCCCAGAGGGACTTCTGAAGCATGCCCCATGATTCTTGAAATCATG  855
          |||                  ||||:::
Db    285 ArgIleIleGlyAspProTyrLysValGlnGlnAlaCysGluMetAspIleLeu  304

Qy    856 CAGAAAGAGGCAGATGAG--------------------------ACCAAACTAGCC  885
          :::|||                                       ::
Db    305 ---ArgGluArgAspGlnGlyGlyPheGlyGlyArgArgAsnGluTyrGlySerArgValGly  323

Qy    886 GAAGAGATTCCTCTGAAAATCTGGCACCACAATGGCTTGGTTGGTGGAGACTGATTGGAAAA  945
          |||||||||                   :::||:
Db    324 GlyGlyIleLeuAspSerValProValProArgHisSer-----ValGlyValValIleGlyArg  341

Qy    946 GAAGGCCAGAAATTGAAGAAAAATTGAACATGAAACATGGGACCAAGATAACAATCCATCT  1005
          |||||||              :::|||||||
Db    342 SerGlyGluMetIleLysLysIleGlnAsnAspArgAlaGlyValAlaArgIleGlnPheLys--  360

Qy   1006 TTGCAGATTTGAGCATATACAACCCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAG  1065
          |||||                    |||
Db    361 ---GlnAspProGlyThr---GlyProGluLysIleAlaHisIleMetGlyProProAsp  378

Qy   1066 GCCTGTGCCAGTGCTGAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGAT  1125
          |||||||                                       ||||||
Db    379 ArgCysGluHisAlaAlaAlaArgIleIle---------------------AsnAsp  389

Qy   1126 ATGCTGGCTGTGTTAACCAACAGCCAATCTGATCCCAGGGTTGAACCTGCCACTTGGC  1185
          |||                                           -----
Db    390 LeuLeuGln------------------------------------------  392

Qy   1186 ATCTTTTCAACAGGACTGTCCGTGCTATCCTGCCACCAGCAGGCCCCGGAGCTCCCCCC  1245
          ||||||              ||||:::
Db    393 ---------------------SerLeuArgSerGlyTyrProProGlyProGlyAlaProGly  406

Qy   1246 GCTGCCCCCACCACCCCTTCACTACCCACCTCCGGATACTTCTCCAGCTGTGACCCCCAT  1305
          |||||                                        -------GlyLyArgGlyArgGlyArgGlyGln  418
Db    407 MetProPro--------------------------------------------------  407

Qy   1306 CACCAGTTGGCCCGGTTCCCGCATCATCACTCTTATCCAGACGACAGGAGATTGTGAATCTC  1365
          |||||                                         :::||||||:::
Db    419 GlyAsnTrpGlyPro-------------------------ProGlyGlyGlu------MetThrPhe  430

Qy   1366 TTCATCCCAACCCAGGCGTGTGGCGCCCATCATCGGCGGAAGAAGGGGCCACACATCAAACAG  1425
          |||||:::                                      ::|||||
Db    431 SerIleProThrHisLysCysGlyLeuValIleLeuGlyArgGlyGlyGluAsnValLysLysAla  450

Qy   1426 CTGGCGAGATTCGCCGGAGCCTCTATCAAGATTGCC------------CCTGCGGAAGGCCCCA  1476
          :::                                              |||
```

```
Db        451  IleAsnGlnGlnThrGlyAlaPheValGluIleSerArgGlnLeuProProAsnGly--  469

Qy       1477  GACGTCACGAAAGATGGTCATCATCACCGGGCACCGGAAGCCAGTTCAAGGCCCAG 1536
                |||  ::: :::::: ::|||| |||  |||   ::::::     |||:||
Db        470  AspProAsnPheLysLeuPheValIleArgGlySerProGlnIleAspHisAlaLys  489

Qy       1537  GGACGGATCTTTGGGAAACTGAAAGAGGAGAAAACTTCTTTAACCCCAAAGAAGAAGTGAAG 1596
                                                              ||| |||
Db        490  GlnLeuIle--------------------------------------GluGluLys  495

Qy       1597  CTGGAAGCGGCATATCAGAGTGCGCCTTCCACAGCTGGCCGCGGGTGATTGGCAAGGTGGC 1656
                ::||||  |||
Db        496  IleGluGly----------------ProLeuCysProValGlyPro----GlyProGlyGly  509

Qy       1657  AAGACCGTGAACGAACGACTGCAGAACTTAACCAGTGAGAAGTCATCGTGCCTCGTGACCAA 1716
                                                                  |||
Db        510                                          ------ProGlyProAla  513

Qy       1717  ACGCCAGATGAAAATGAGGAAGTCAGATTATCGGGCACTTCTTTGCTAGCCAG 1776
                                                              ::|||||
Db        514  GlyPro--------------------------MetGlyProPheHisProGlyPro  523

Qy       1777  ACTGCACAGCGCAAGATCAGGAAATTGTACAACAGGTGAAGCAGCAGCAGAAATAC 1836
                                                              |||
Db        524  PheAsnGlnGly--------------------------------------------  527

Qy       1837  CCTCAGGGAGTCGCCTCACAGCGGCAGCAAGTGAG-------------GCTCCCACAGG 1881
                |||  :::||||       ||              ||
Db        528  ProProGly-AlaProHisAlaGlyGlyProProProHisGlnTyrProProGlnGl  547

Qy       1882  CACCAGCAAAACAACGGATGAATGTAGCCCTTCCAACACCTGACAGAATGAGACCAAACG 1941
                |||                ::||| |||     |||||
Db        547  yTrpGlyAsnThrTyrProGluTrpGlnPro-ProAlaProHis------AspProAsnL  565

Qy       1942  CAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAGA--------ATGAGAAGTC 1992
                ||||||  :::   ||   |||           ||
Db        565  ysAlaAlaAlaAlaThrAspProAsnAlaAlaTrpAlaAlaTyrTyrSerHisTyrT 585

Qy       1993  TGCGGAGGCGGCCAGGGACTCTGCCGAGGACCCTGGGGCCGAGGAGGGCG 2052
                :::||||||             ::|||
Db        585  yrGlnGlnProProGlyProValPro--------------------GlyProAlaProAlaP 599

Qy       2053  GGGAAGGTCAGCCAGGTTGCCAGAACCACCGAGCCCCGCTCCCGCCCCCCCAG 2106
                |||||||||||        |||||||||
Db        599  roAlaAlaProAlaGlnGlnProGlnProProGln-ProProProThrGlyGln  616
```

RESULT 2
```
FUB3_HUMAN
ID   FUB3_HUMAN     STANDARD;     PRT;    572 AA.
AC   Q96I24; Q92946; Q9BVB6;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Far upstream element binding protein 3 (FUSE binding protein 3).
GN   FUB3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Cervix, and Placenta;
RX   MEDLINE=97094955; PubMed=8940189;
RA   Davis-Smyth T., Duncan R.C., Zheng T., Michelotti G., Levens D.;
RT   "The far upstream element-binding proteins comprise an ancient family
RT   of single-strand DNA-binding transactivators.";
RL   J. Biol. Chem. 271:31679-31687(1996).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   TISSUE=Cervix, and Placenta.
RX   MEDLINE=23388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: May interact with single-stranded DNA from the far-
CC       upstream element (FUSE). May activate gene expression.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q96I24-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q96I24-2; Sequence=VSP_008323, VSP_008324;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Detected in a number of cell lines.
CC   -!- SIMILARITY: Contains 4 KH domains.
CC   -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC       frameshift in position 18.
CC   -!- CAUTION: Ref.2 (AAH01325) sequence differs from that shown due to
CC       a frameshift in position 527.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U69127; AAC50893.1; ALT_FRAME.
DR   EMBL; BC001325; AAH01325.1; ALT_FRAME.
DR   EMBL; BC007874; AAH07874.1; -.
DR   Genew; HGNC:4005; FUBP3.
DR   MIM; 603536; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
KW   Transcription regulation; Trans-acting factor; Nuclear protein;
KW   DNA-binding; Repeat; Alternative splicing.
FT   DOMAIN        77    141       KH 1.
FT   DOMAIN       162    228       KH 2.
FT   DOMAIN       253    317       KH 3.
FT   DOMAIN       354    421       KH 4.
FT   VARSPLIC       1     64       MAELVQGQSAPVGMKAEGFVDALHRVRQIAAKIDSIPHLNN
FT                                 STPLVDPSVYGYGVGVQKRPLDDGV -> MPPI (in
FT                                 isoform 2).
FT                                 /FTId=VSP_008323.
FT   VARSPLIC     298    548       Missing (in isoform 2).
FT                                 /FTId=VSP_008324.
FT   CONFLICT      72     72       V -> D (IN REF. 1).
SQ   SEQUENCE   572 AA;  61640 MW;  F1BE225428C197D CRC64;
```

Alignment Scores:
```
Pred. No.:                                           572
Score:               4.06e-09      Length:            572
Percent Similarity:  262.50        Matches:           119
Best Local Similarity: 37.43%      Conservative:       82
Query Match:          22.16%       Mismatches:        204
                       4.29%       Indels:            133
```

```
DB:                    1              Gaps:          17

US-09-270-437D-6 (1-3412) x FUB3_HUMAN (1-572)

Qy    496 CATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGATGAGAGGTGAGCTCC 555
          ||||||:||| :::: |||||||
Db     24 HisArgValArgGlnIleAlaIleAlaLysIleAspSerIlePro-----HisLeuAsnAsn 41

Qy    556 CCTTCGCCCCCTCAGCCGAGCCCAGCGTG------GGGACCACTCCTTCC 597
          ::: |||||| :||||||||||
Db     42 SerThr-ProLeuValAspProSerValTyrGlyTyrGlyValGlnLysArgProLeuAs 61

Qy    598 CGGGAGCAAGGCCACGCCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTG 657
          |||||||||::: ::|||||||
Db     61 pAspGlyValGlyAsnGlnLeuGlyAlaLeuValHisGlnArgThrValIleThrGlu-- 80

Qy    658 CGGATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATA 717
          |||:|||||| ::: |||||
Db     81 GluPheLysValProAspLysMetValGlyPheIleIleGlyArgGlyGlyGluGlnIle 100

Qy    718 AAGAACATCACTAAGCACAGACCCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGA 777
          ||| :::::: ::||:|| ||| ::||||||||
Db    101 SerArgIleGlnAlaGluSerGlyCysIleLysIleGlnIle--AlaSerGluSerGly 119

Qy    778 GCTGCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGC 837
          :: :|| ::|||||||
Db    120 IleProGluProCysValLeuThrGlyProGluSerIleGluGlnIleAlaLysArg 139

Qy    838 ATGATTCTTGAAATCATGCAGAAA------GAGGCAGAT 870
          :::::: |||| ::|||||
Db    140 LeuLeuGlyIleLeuValAspArgCysArgAsnGlyProGlyPheHisAsnAspIleAsp 159

Qy    871 GAGACCAAACTAGCCGAAGAGATTCCTCTGAAAATCTTGGCACACAAATGGCTTGGTTGGA 930
          ::||||||
Db    160 SerAsnSerThrIleGlnGlu-------IleLeuIleProAlaSerLysValGly 175

Qy    931 AGACTGATTGGAAAAGAAGGCAGAAAATTGAACAAAATTGAACACAGGGACCAAG 990
          ::|||:: |||| :::|||||||
Db    176 LeuValIleGlyArgGlyGlyGluThrIleLysGlnLeuGlnLeuLysGlyValLys 195

Qy    991 ATAACAATCTCATCTTGCAGGATTTGAGCATA--TACAACCCGGAAAGAACCATCACT 1047
          ::: ::|||||||:: :::||
Db    196 MetValMet------IleGlnAspGlyProThrGlyAlaAspLysProLeuArg 213

Qy   1048 GTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGAGATAGAGATTATGAAGAAGCTGCGT 1107
          ||| :||||||::: ::|||
Db    214 IleThrGlyAspProAlaPheLysValGlnIleAlaIleArgGluMetValLeuGlulleIleArg 233

Qy   1108 GAGGCCTTTGAAAATGATATGCTGGCTGTGTTAACCAACAAGCCAATCTCATCCCCAGGGTTG 1167
          |||
Db    234 GluLysAspGlnIleAlaAspPhe------------ 240

Qy   1168 AACCTCAGGCGACACTTGGCATCTTTTCAAGACTGTCCGTGCTATCTCCACCAGCAGGG 1227
          |||
Db    241 ------------------------ArgGly 242

Qy   1228 CCCCGCGGAGCTCCCCCGCTGCCCCTTCACTACCACCCCCTTCACTGAGATTCCGGATACTTC 1287
          :: :|||||
Db    243 ValArgGly-----------------AspPheAsnSerArgMetGlyGlyGly 254

Qy   1288 TCCAGCCTGTACCCCCATCCAGTTGGCCGTTCCGCCGTTCCCCATCATCACTCTTATCCAGAG 1347
          |||
Db    255 Ser------------------------- 255

Qy   1348 CAGGAGATTGTGAATCTTCATCCCAACCCAGGCTGTGGGGCCCATCATCGGGAAGAAG 1407
          |||||||||:::|||:::
Db    256 --------IleGluValIleSerValProArgPheAlaValGlyIlelevalIleGlyArgAsn 272

Qy   1408 GGGGCCACCATCAAACAGTCTGGCGAGATTCGCCGGAGCCTCTATCAAGATTGCCCCTGCG 1467
          |||||:||:::|||||:::
Db    273 GlyGluMetIleLysLysGlnLeuGlnAsnAspAlaGlyValArgIleGlnProLysAsp 292

Qy   1468 GAAGGCCCAGACGTCAGCGAAGATGGTCATCATCACCGGGCCACCGGAAGCCCAGTTC 1527
```

```
Db    293 AspGlyIleSer---ProGluArgAlaAlaAlaGlnValMetGlyProProAsp--ArgCys 310
          :::|||                                    |||:||::|::   :::
Qy   1528 AAGGCCCAGGACGGACGGATCTTTGGGAAACTG----------AAAGAGGAAAACTTCTTT 1575
          :::                          |||  :::|||
Db    311 GlnHisAlaAlaHisIleIleSerGluLeuThrAlaGlnLeuLeuArgAspGlyAspPhe 330

Qy   1576 AACCCCAAAGAAGAGTGAAG----------------------------------- 1596
          |||||
Db    331 GlyGlyLeuAlaLeuAlaLeuArgGlyArgArgGlyAspTrpSerValGlyValAla 350

Qy   1597 ----------CTGGAGGCCATATCAGAGTGCCCTCTTCCACAGTGCCTGGCCGGTGATT 1644
          |||||||||||                  :::||:|||||
Db    351 ProGlyGlyValGlnGlyIleLeuThrTyrThrValProAlaAspLysCysGlyLeuVallle 370

Qy   1645 GGCAAAGGTGGCAAGACCGTGAACGAACTGCAGAACTTAACCAGTGCAGAAGTCATCGTG 1704
          |||||||||||::: :::                      ::: ||||||
Db    371 GlyLysGlyGlyLysAsnIleLysSerGlyAlaHissValGlnLeuLeu 390

Qy   1705 CCTCGTGACCAAAACGCCAGATGAAAATGAGAAGTGATCGTCAGAATTATCGGGCACTTC 1764
          :|||:::                                         :::
Db    391 GlnAsnArgAsnProProAsnSerAspProAsnLeu----------- 402

Qy   1765 TTTGCTAGCCAGACTGCACAGCCAGATCAGGAGAAATTGTACACAGGTGAAGCAGCAG 1824
          :::                        ::: :::|||||:::||||
Db    403 --------------ArgArgPheThrIleArgGlyValProGlnIleIleGluValAla 417

Qy   1825 GAGCAGAAATACCC----------TCAGGAGTCGCCTCACAGCCGCAG- 1862
          ||||||                 ||||||||||
Db    418 ArgGlnIleLeuIleAspGluLysValGlyLyThrAsnLeuGlyAlaProGlyAlaPheGly 437

Qy   1863 CAAGTGGAGGCTCCCACCAGGACCAGCAGCAAAACGATGAAGTAGCCCTTCCACACACT 1922
          |||                                           |||
Db    438 GlnSerProPheSerGlnProProAlaProProHisGlnAsnThrPheProProArgSer 457

Qy   1923 GACAGAATGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCA 1967
          |||||||                                  |||
Db    458 SerGlyCysPheProAsnMetAlaAlaLysValAsnGlyAsnPro 472
```

RESULT 3

ZBP2_CHICK

ID   ZBP2_CHICK   STANDARD;   PRT;   769 AA.
AC   Q8UVD9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zipcode-binding protein 2.
GN   ZBP2.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A., SEQUENCE OF 104-119; 242-267; 679-685; 705-711 AND
RP   743-766, FUNCTION, DEVELOPMENTAL STAGE, AND SUBCELLULAR LOCATION.
RX   MEDLINE=21640447; PubMed=11781334;
RA   Gu W., Pan F., Zhang H., Bassell G.J., Singer R.H.;
RT   "A predominantly nuclear protein affecting cytoplasmic localization of
RT   beta-actin mRNA in fibroblasts and neurons.";
RL   J. Cell Biol. 156:41-52(2002).
CC   -!- FUNCTION: Binds to a 54-nucleotide localization signal (the
CC   zipcode) found in the 3' untranslated region of beta-actin mRNA
CC   and may play a role in its localization.
CC   -!- SUBCELLULAR LOCATION: Predominantly nuclear. Also detected in the
CC   cytoplasm.
CC   -!- DEVELOPMENTAL STAGE: The highest expression is found in 6-d
CC   embryos, is reduced to 30% before hatching and remains stable
CC   thereafter.
CC   -!- SIMILARITY: Contains 4 KH domains.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -

```
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AF461020; AAL66365.1; -.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0003730; F:mRNA 3' UTR binding; IDA.
DR   GO; GO:0008298; P:mRNA localization, intracellular; NAS.
DR   InterPro; IPR004087; KH dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 4.
DR   SMART; SM00322; KH; 4.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
KW   Transport; mRNA transport; Nuclear protein; RNA-binding; Repeat.
FT   DOMAIN     208    272       KH 1.
FT   DOMAIN     296    362       KH 2.
FT   DOMAIN     385    449       KH 3.
FT   DOMAIN     486    553       KH 4.
SQ   SEQUENCE   769 AA;  80644 MW;  875A6C83529969EB CRC64;
```

```
Alignment Scores:
Pred. No.:                  5.89e-09        Length:            769
Score:                      260.50          Matches:           142
Percent Similarity:         36.0%           Conservative:       67
Best Local Similarity:      24.48%          Mismatches:        173
Query Match:                4.26%           Indels:            198
Db:                         1               Gaps:               31
```

US-09-270-437D-6 (1-3412) x ZBP2_CHICK (1-769)

```
Qy      12 AGGCGAGGAGCGCCGGGTACCGGCCGGGGGAGCCCTCTCGGGGGAAGAGACCGGAT 71
            ||||||                    |||||||||||  ||||
Db     160 ArgSerGlyProValGlyAspProGlyProProGlyProProArgAla----------- 173

Qy      72 GATGAACAAGCTTTACATGGGACCTGAGCCCTCCGGCCCCGCGACCTCCGGCA 131
                                     |||||||||||
Db     174 --------------------------GluArgGlyArgProProGlyProPro 181

Qy     132 GCTCCTTTGGGGACAGAAGCTGCCCCTGGCGGGACAGGTCCTGCTGAAGTCCGGCTACGC 191
            ||||||                    ||||
Db     182 AlaLeu---------------------ThrGlyGlyAlaLeuProSer----------- 190

Qy     192 CTTCGTGGACTACCCCGACCAGAACTGGGC------CATCCG----------- 227
            |||||||       |||::|||||         |||::||
Db     191 --AlaAlaLeuPro---ProGlnLeuGlyProMetHisProProArgSerThrThr 208

Qy     228 -----------CGCCATCGAGACCCCTCTGCGGGTAAAGTGCATGGGAAAAATCATGGA 278
                        :::||||                 ||||||  ||
Db     209 ValThrGluGluTyrArgValProAspGlyMetValGlyLeuIleIleGlyArgGlyGly 228

Qy     279 AGTTGATTACTTCAGTGTCTCTAAAAAGCTAAGGAGCAGGAAAATTCAGATTCGAAACATCCC 338
            ||:                    ||||||::|||||
Db     229 GluGlnIle-------------------AsnLysIleGlnAspSerGlyCysLysValGln 243

Qy     339 TCCTCACCTGCAGTGGCGGAGCGTGTTGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAA 398
            |||||||       ||||::::|             ||::||
Db     244 IleSerProSerGlyLeu----------------- 251

Qy     399 TGTGAACAAGCAAAAATAGCCATGAGAGAAACCGCCGTTGTCAAGTCACATATGCAACAAG 458
                                  ::|||  ||||||       |||  :||  ||
Db     252 ---------------------------------------------------------Pro 252

Qy     459 AGAAGAAGCAAAAATAGCCATGAGAGAAGCTAAGCGGGATTCAGTTGAGAAACTTACTCCTT 518
                                  ::||       ||||  |||       :::|||
Db     253 GluArgGlySerValSerLeuThrGlyValProAspGlyAspGlyAspGlyAspGlyGlyPro 272

Qy     519 CAAGATTTCCTACATCCCGATGAGAAGAGGTGAGCTCCCCTTCGCCCCCCTCAGCGAGCCCA 578
            ||||||||
Db     273 AspAspIleValSerArgGly--ArgGly----------------------- 281
```

```
Qy     579 GGGTGGGGACCACTCTTCCCGGGAGCAGGCCAC----------GCCCCTGGGGCCAC 626
            |||||||                              |||||
Db     282 -----------GlyPro-----ProGly--GlnPheHisAspTyrAlaAsnGlyGlnAsnGlyTh 297

Qy     627 TTCTCAGGCCAGACAGATTGATTTCCCGCTGCGGATCCTGGTCCCCACCAGTTTGTTGG 686
            ||
Db     297 rValGln-----------------GluIleMetIleProAlaGlyLysLysAlaGl 309

Qy     687 TGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAACATCACTAAGCAGCACCAGTCCG 746
            |||::|||    |||||||||       |::||||||
Db     309 yLeuValIleGlyLysGlyGlyGluThrIleLysGlnGlnLeuLysGlnAlaGlyValLy 329

Qy     747 G--GTAGATATCCATAGAAAAGAGAACTCTGGAGCTGCAGAGAAGAAGCCTGTCACCATCA 803
            :  :::||||             ::|||||||:::
Db     329 sMetIlePheIleGlnAspGlySerGlnAsnThrAsnValAspLysProLeuArgIleIl 349

Qy     804 TGCCACCCCAGAGAGGGACTTCTGAAGCATGCGCATGATTCTTGAAATCATGCAGAAAGA 863
            ||       |||     |||||    |||::||    :::|||||||||::|
Db     349 eGlyAspProTyrLysValGlnGlnAlaCysGluMetValMetAspIleLeu---ArgGl 368

Qy     864 GGGCAGATGAG----------------ACCAAACTAGCCGAAGAGAT 893
            ||||||                   ::::::
Db     368 uArgAspGlnGlyGlyPheGlyAspArgAsnGluTyrGlySerArgIleGlyGlyGlyIl 388

Qy     894 TCCTCTGAAAATCTTGGCACACACAATGGCTTGGTTGGAAGACTGATTGGAAAGAAGGCAG 953
            :::||||       ||||::|        |||||||||
Db     388 eAspValProValProArgHisSer-----ValGlyValValIleGlyArgSerGlyGl 406

Qy     954 AAATTTGAAGAAAATTGAACATGAAACAGGACCAAGATAACAATCTCATCTTTGCAGGA 1013
            :::||||       ||||||||::||
Db     406 uMetIleLysLysGlnIleGlnAsnAspAlaGlyValArgIleGlnPheLys------GlnAs 424

Qy    1014 TTTGAGCATATACAACCCGGAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGC 1073
            :::||||                  :::|||
Db     424 pAspGlyThr--GlyProGluGluLysIleAlaHisIleMetGlyProGluArgCysGl 443

Qy    1074 CAGTGCTGAGATAGAGATTATGAAGAGCTGCGTGAGGCCTTTGAAAATGATGATGCTGGC 1133
            ||       :::|||    ||||
Db     443 uHisAlaAlaArgIleIle------------------AsnAspLeuLeuGl 454

Qy    1134 TGTTAACCAACAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTC 1193
            :::|||||||        ||||  |||||:::
Db     454 n-------------------------------------------------------- 454

Qy    1194 AACAGGACTGTCCGTGCTATCTCCACCAGCAGGGCC------CGCGAGCTCCCCCGC 1247
            |||::|||       |||||  |           |||
Db     455 -----------SerLeuArgSerGlyProProProGlyHisGlyMetProProProGl 471

Qy    1248 TGCCCCCTACCACCCCTTCACTACCCACTCCGATACTTCTCCAGCTCTGTACCCCATCA 1307
            ||       :::|||                   |||
Db     471 yGly-----------------------ArgGlyArgGlyGlnGlyIleTrp--------- 482

Qy    1308 CCAGTTGGCCCGTTCCCGCCATCATCACTTCATTCCAGAGCAGGAGATTGTGAATCTCTT 1367
            ||       |||                 ||
Db     483 --------GlyPro-----------ProGlyGlyGlu----MetThrPheSe 492

Qy    1368 CATCCCCAACCCAGGCTGTGGGCGCCATCATCGGGAAGAAGGGGCACACATCAAACAGCT 1427
            |||||||       ||:::|             ::|||||
Db     492 rIleProThrHisLysCysGlyLeuValIleIleGlyArgGlyGlyGluAsnValLysAlaiI 512

Qy    1428 GGGGAGATTCGGCCAGGGCTTCTATCAAGATTGCC----------CCTGCGGAAGGCCCAGA 1478
            |||:::|                            |||||
Db     512 eAsnGlnLeuArgGlyValAlaPheValGluIleSerArgGlnLeuProProAsnGly---As 531

Qy    1479 CGTCAGCGAAAGATGGTGTCATCATCACCGGGCACCGGAAGCCCAGTTCAAGGCCCAGGG 1538
            :::||       :::||||||||
Db     531 pProAsnPheLysLeuPheIleIleArgGlySerProGln----------------- 544

Qy    1539 ACGGATCTTTGGGAAACTAAGGAGAGAAAACTTCTTTAACCCCAAAGAAGGAAGTGAAGCT 1598
            |||
Db     545 -----------------GlnIleGluHisAlaLysGlnProIleGluGlu--LysIl 557

Qy    1599 GGAAGCGCATATCAGAGTGCCCTCTTCCACAGCTGGCCGGTGATTGGCAAAGGTGGC 1656
```

```
Db    557 eGluGly----------ProLeuCysProValGlyProGlyProGlyProGlyGly 572
              :|||              |||              |||   |||||
```

RESULT 4
```
ID   FUB2_HUMAN    STANDARD;    PRT;   707 AA.
AC   Q92945; O00301; Q9UNT5; Q9UQH5;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Far upstream element binding protein 2 (FUSE binding protein 2) (KH
DE   type splicing regulatory protein) (KSRP) (p75).
GN   FUBP2 OR KHSRP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., SEQUENCE OF 72-85; 123-128; 267-281; 283-291;
RP   348-359; 472-486; 488-492; 620-625; 627-644 AND 646-651, AND FUNCTION.
RC   TISSUE=Neuroblastoma, and Retinoblastoma;
RX   MEDLINE=97282621; PubMed=9136930;
RA   Min H., Turck C.W., Nikolic J.M., Black D.L.;
RT   "A new regulatory protein, KSRP, mediates exon inclusion through an
RT   intronic splicing enhancer.";
RL   Genes Dev. 11:1023-1036(1997).
RN   [2]
RP   SEQUENCE OF 1-115 AND 570-707 FROM N.A.
RX   MEDLINE=9189245; PubMed=8940189;
RA   Ring H.Z., Vameghi-Meyers V., Nikolic J.M., Min H., Black D.L.,
RA   Francke U.;
RT   "Mapping of the KSRP gene to a region of conserved synteny on human
RT   chromosome 19p13.3 and mouse chromosome 17.";
RL   Genomics 56:350-352(1999).
RN   [3]
RP   SEQUENCE OF 60-707 FROM N.A., AND FUNCTION.
RC   TISSUE=B-cell lymphoma, and Skeletal muscle;
RX   MEDLINE=97094955; PubMed=8940189;
RA   Davis-Smyth T., Duncan R.C., Zheng T., Michelotti G., Levens D.;
RT   "The far upstream element-binding proteins comprise an ancient family
RT   of single-strand DNA-binding transactivators.";
RL   J. Biol. Chem. 271:31679-31687(1996).
RN   [4]
RP   FUNCTION, AND INTERACTION WITH PTBP1; PTBP2 AND HNRPH1.
RX   MEDLINE=20459250; PubMed=11003644;
RA   Markovtsov V., Nikolic J.M., Goldman J.A., Turck C.W., Chou M.-Y.,
RA   Black D.L.;
RT   "Cooperative assembly of an hnRNP complex induced by a
RT   tissue-specific homolog of polypyrimidine tract binding protein.";
RL   Mol. Cell. Biol. 20:7463-7479(2000).
CC   -!- FUNCTION: Binds to the dendritic targeting element and may play a
CC       role in mRNA trafficking (By similarity). Part of a ternary
CC       complex that binds to the downstream control sequence (DCS) of the
CC       pre-mRNA. Mediates exon inclusion in transcripts that are subject
CC       to tissue-specific alternative splicing. May interact with single-
CC       stranded DNA from the far-upstream element (FUSE). May activate
CC       gene expression.
CC   -!- SUBUNIT: Part of a ternary complex containing FUBP2, PTBP1, PTBP2
CC       and HNRPH1.
CC   -!- SUBCELLULAR LOCATION: Nuclear. A small proportion is also found in
CC       the cytoplasm of neuronal cell bodies and dendrites (By
CC       similarity).
CC   -!- TISSUE SPECIFICITY: Detected in neural and non-neural cell lines.
CC   -!- SIMILARITY: Contains 4 KH domains.
CC   -!- CAUTION: Ref.3 sequence differs from that shown due to numerous
CC       frameshifts.
```

```
DR   EMBL; U94832; AAB53222.1; -.
DR   EMBL; AF093747; AAD29861.1; -.
DR   EMBL; AF093745; AAD29861.1; JOINED.
DR   EMBL; AF093748; AAD29862.1; -.
DR   EMBL; U69126; AAC50892.1; ALT_FRAME.
DR   Genew; HGNC:6316; KHSRP.
DR   MIM; 603445; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 4.
DR   SMART; SM00322; KH; 4.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
KW   Transport; mRNA transport; mRNA processing; mRNA splicing;
KW   Transcription regulation; Trans-acting factor; Nuclear protein;
KW   DNA-binding; RNA-binding; Repeat.
FT   DOMAIN      144    208       KH 1.
FT   DOMAIN      233    299       KH 2.
FT   DOMAIN      322    386       KH 3.
FT   DOMAIN      422    489       KH 4.
FT   DOMAIN      569    682       4 X 12 AA IMPERFECT REPEATS.
FT   REPEAT      569    580       1.
FT   REPEAT      615    626       2.
FT   REPEAT      641    652       3.
FT   REPEAT      671    682       4.
FT   DOMAIN        7     67       GLY/PRO-RICH.
FT   DOMAIN       68    495       GLY-RICH.
FT   DOMAIN      496    610       ALA/GLY/PRO-RICH.
FT   CONFLICT     96     96       G -> V (IN REF. 2).
FT   CONFLICT    692    692       V -> G (IN REF. 2).
FT   CONFLICT    694    694       G -> GG (IN REF. 2).
FT   CONFLICT    707    707       Q -> A (IN REF. 2).
SQ   SEQUENCE    707 AA;  72709 MW;  E07588DE43BCA8B6 CRC64;
```

Alignment Scores:
```
Pred. No.:             2.32e-08      Length:        707
Score:                 251.00       Matches:       131
Percent Similarity:    35.95%       Conservative:   75
Best Local Similarity: 22.86%       Mismatches:    181
Query Match:           4.10%        Indels:        186
DB:                    1            Gaps:           27
```

US-09-270-437D-6 (1-3412) x FUB2_HUMAN (1-707)

```
Qy   35 GCCGGGGAGCCGGCCGGCTCCTGGGGAGGAGAGACGGATGATGAACAAGCTTTACATGGGA 94
           ||||||||||||| ::|        |||
Db   17 AlaGlyGlyGlyGlyGlyGlyAlaGlyGly--------------------------AlaGly 27

Qy   95 ACCTGAGCCCCGCCGTCACCGCCGACC---------------------------------- 124
Db   28 GlyGlyProProGlyProProGlyAlaGlyAspArgGlyGlyGlyProCysGly 47

Qy  125 ----------TCCGGCAGCTCTTTGGGAGCACAGGAGCTGCCCCTGGCCGGGGACCAGGTCC 172
                  |||||::: :||  ||||
Db   48 GlyGlyProGlyGlyGlySerAla-GlyGlyProSerGlnProGlyGlyGlyGlyPr 67

Qy  173 TGCTGAAGTCCGGCTACGCCTTCGTGGACTACCCCGACCAGAACTGGGCCATCCGCGCCA 232
           ||||    |||
Db   67 OGlyIleArgGlys-AspAlaPheAlaAspAlaVaLGlnArgAlaArgGlnIleAlaAlaAla- 86

Qy  233 TCGAGACCCTCTCCGGGTAAAGTGGGATTGCATGGG----------------------- 267
           :::
Db   87 ----LysIleGlyGlyAspAlaAlaAlaThrThrGlyAsnAsnSerThrProAspPheGlyP 105

Qy  268 -------------------AAAATCATGGAAGTTGATTACTACCAGTCTCTAAAAGCTAAGGA 310
Db  105 heGlyGlyGlnIleuGlysArgGlnIeuGlusAspGlyAspGlnProGlusSerLysGlysLeuAlaS 125

Qy  311 GCAGGAAAATTCAGATTCGA------------------AACATCCCTCCTCCACCTG---- 348
           :::                               |||
Db  125 erGlnGlyAspSerIleSerSerGlnLeuGlyProIleHisProProProArgThrSerM 145
```

Qy   349 ------CAGTGGGGAGGCTGTTGGATGGACTTTTGGCTCAATATGGGACAGTG------ 393
                  ::||||| ::|||||
Db   419 rpGlyGlyProGlyGlyGlu--------------------------MetThrPheSerIleP 430

Db   145 etThrGluGluTyrArgValProAspGlyMetVal--------GlyLeuIleIleGlyA 162
Qy   1373 CAACCCAGGCTGTGGGCGCCATCATGGGAAGAAGAAGGGGCACACATCAAACAGCTGGCGA 1432

Qy   394 ------GAGAATGTGGAACAAGTCAACACAGAACACGCCGTTGTCAACGTCA------ 445
                  ||: :::|||| :::||||| on
Db   430 roThrHisLysCysGlyLeuValIleGlyLysGlyGlyGluAsnValLysAlaIleAsnG 450

Db   162 rgGlyGlyGluGluIleAsnLysIleGlnGlnAspSer------ 174
Qy   1433 GATTCGCCGGAGCCTCTATCAAGATTGCC------CCTCGGGAAGGCCCAGAGCGTCA 1483

Qy   446 CATATGCAACAAGAGAAGAAGCAAAAATAGCCATGGAGAGCTAAGCGGGCATCAGTTG 505
                  ::: |||||||:::   :::|||:||
Db   175 ------GlyCysLysLysValGlnIleSerProAspSerGlyGly------ 186

Db   186 lnGlnThrGlyAlaPheValGluIleSerArgGlnLeuProProThrGlyThrProThrS 470

Qy   506 AGAACTACTCCTTCAAGATTCCTACATCCCGGATGAAGAGGTGAGCTCCCCTTCGCCCC 565
                  :::|||
Qy   1484 GCGAAAGGATGGTCATCATCACCGGCGCCACCGAA 1518

Db   187 ------LeuProGluArgSerValSerLeuThrGlyAlaP 198
Db   470 er---LysPheIleIleArgGlySerProGln 480

Qy   566 CTCAGGCGAGCCCAGCCTGGG------GACCACTCTTCCCGGGAGCGAGGCC 610

RESULT 5

ID   PCB3_HUMAN   STANDARD;   PRT;   339 AA.
AC   P57721;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Poly(rC)-binding protein 3 (Alpha-CP3).
GN   PCB3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20396135; PubMed=10936052;
RA   Makeyev A.V., Liebhaber S.A.;
RT   "Identification of two novel mammalian genes establishes a subfamily
RT   of KH-domain RNA-binding proteins.";
RL   Genomics 67:301-316(2000).
CC   -!- FUNCTION: Single-stranded nucleic acid binding protein that binds
CC   preferentially to oligo dC (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: Contains 3 KH domains.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; AF176329; AAG09240.1; -.
DR   HSSP; Q07244; 1KHM.
DR   Genew; HGNC:8651; PCBP3.
DR   GO; GO:0005634; C:nucleus; ISS.
DR   GO; GO:0003677; F:DNA binding; ISS.
DR   GO; GO:0003723; F:RNA binding; NAS.
DR   GO; GO:0016071; P:mRNA metabolism; NAS.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding;
KW   Repeat.
FT   DOMAIN   13   63   KH 1.
FT   DOMAIN   97   150   KH 2.
FT   DOMAIN   261   313   KH 3.
SQ   SEQUENCE   339 AA; 35938 MW; F9F19FF7590C4188 CRC64;

Alignment Scores:

Pred. No.:		
Score:	4.27e-08	Length: 339
Percent Similarity:	245.50	Matches: 90
Best Local Similarity:	38.65%	Conservative: 53
Query Match:	24.32%	Mismatches: 116
	4.01%	Indels: 111

Db   198 roGluSerValGlnLeuAlaLysMetMetLeuAspAspIleValSerArgGlyArgGlyG 218

Qy   611 ACGCCCTGG------GGCACTTCTCAGGCCA 637

Db   218 lyProProGlyGlnPheHisAspAsnAlaAsnGlyGlyGlnAsnGlyThrValGln-- 236

Qy   638 GACAGATTGATTTCCCGCTGCGGATCCTGTCCCCACCCAGTTGTTGTGTGCCATCATCG 697

Db   237 ------GluIleMetIleProAlaGlyLysLysAlaGlyLeuValIleG 250

Qy   698 GAAAGGAGGCGTTGACCATAAAGAACATCACTAAGCAACAGACCCCAGTCCCGGGTAGATATCC 757

Db   250 lyLysLysGlyGluThrIle--------LysGlnLeuGlnGluArgAlaGlyValL 266

Qy   758 ATAGA------AAAGAACTCTGGAGCTGCAGAGAGCCCTGTCACCA 799

Db   266 ysMetIleLeuIleGlnAspGlySerGlnAsnThrAsn---ValAspLysProLeuArgI 285

Qy   800 TCCATGCCACCCCAGAGGGGACTTCTGAGCATGCGCATGATTCTTGAAATCATGCAGA 859

Db   285 leIleGlyAspProTyrLysValGlnIleAlaCysGluMetValMetAspIleLeuArgA 305

Qy   860 AA------ACCAAACTAGCCGAAGAGA 892

Db   305 snValThrLysAlaGlyPheGlyAspArgAsnGluTyrGlySerArgIleGlyGlyIl 325

Qy   893 TTCCTCTGAAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCA 952

Db   325 leAspValProArgHisSer------ValGlyValValIleGlyArgSerGlyYG 343

Qy   953 GAAATTTGAAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCCTCATCTTTGCAGG 1012

Db   343 luMetIleLeuLysIleGlnAsnAspAlaGlyValArgIleGlu-----GlnA 361

Qy   1013 ATTTGAGCATACAACACCCGGAAAGAACCATCACTGAAGGGCACAGTTGAGCCTGTG 1072

Db   361 spAspGlyThr--GlyProGluLysIleAlaHisIleMetGlyProProAspArgCysG 380

Qy   1073 CCAGTGCTCGAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGCTGG 1132

Db   380 ------AsnAspLeuLeuG 391

Qy   1133 CTGTTAACCAACAGCCAATCTGATCCCAGGGTTGAACCTCAGCCACTTGGCATCTTTT 1192

Db   391 ln------ 391

Qy   391 ln------ 391

Qy   1193 CAACAGGACTGTCCGTGCATCTCCACCACCAGGCCCGCGGAGCTCCCCCGCTGCCC 1252

Qy   1253 CCTACCACCCCTTCACTACCCACTCCGGATACTTCTCCCAGCCTCACCCCATCACCAGT 1312

Db   408 lyGlyArgGlyArgGlyArgGlyGlnGly------AsnT 419

Qy   1313 TTGGCCCGTTCCCGACCATCATCACTCTTATCCAGAGCAGGAGAGATTGTGAATCTCTTCATCC 1372

DB:

US-09-270-437D-6 (1-3412) x PCB3_HUMAN (1-339)

```
Qy          1                                                 Gaps:        14

Qy    643  ATTGATTTCCCGCTGCGGATCCTGGTCCCCAGTTGTTGGTGCCATCATCGGAAAG  702
           ::: |||||||:||:|:|::    :::  ||||||:||||||||||||||||||
Db     12  ValThrLeuThrIleArgLeuLeuMetHisGlyLysGluValGlySerIleIleGlyLys  31

Qy    703  GAGGGCTTGACCATAAAGAACATCACTAAGCAGACCCAGTCCCGGTAGATATCCATAGA  762
           ::: |||||||:||    :::  ::::::  ::|:|||||||:||------  ::|
Db     32  LysGlyGluThrValLysMetArgGluValSerGlyAlaArgIleAsnIle------  49

Qy    763  AAAGAGAACTCTGGAGCTGCGAGAGAAGCCTGTCACCATCCATGCACCCCAGAGGGGGACT  822
           ||||||  :::  :::||||||||||||||    :::
Db     50  --SerGlyAsnCysProGluArgIleValThrIleThrGlyProThrAspAlaIle  68

Qy    823  TCTGAAGCATGCCGCCATGATTCTT----------------GAAATCATGCAGAAAGAGGCA  867
           :::|||||    ||||||                 :::|:::::  ::|
Db     69  PheLysAlaPheAlaMetIleAlaTyrLysPheGluGluAspIleIleAsnSerMetSer  88

Qy    868  GATGAGACCAAACTAGCCGAA--GAGATTCCTCTGAAAATCTTGGCACCACAATGGCTTG  924
           ::: ||||||||||||||||   :::
Db     89  AsnSerProAlaThrSerLysProProValThrLeuArgLeuValValProAlaSerGln  108

Qy    925  GTTGAAGACTGATTGGAAAGAAGGCAGAAAATTGAACAAGAAATTGAACATGAAACAGGG  984
           :::|||||||  :::||:||:|:::    ||||||
Db    109  CysGlySerLeuIleGlyLysGlyGlySerLysIleLysGluIleArgGluSerThrGly  128

Qy    985  ACCAAGATAACAATCTCATCTTTGCAGATTGAGCATATACAACCCGGAAAGAACCATC  1044
           ||||||||  ::::  ::|  :::||||||||||||||
Db    129  AlaGlnValGlnValAlaGly------AspMetLeuProAsnSerThrGluArgAlaVal  146

Qy   1045  ACTGTGAAGGCACAGTTGAGGCCTGCCAGTGCTGAGATAGAGATTATGAAGAAGCTG  1104
           ||||||  :::|||  ||||||    ::|||
Db    147  ThrIleSerGlyThrProAspAlaIleIleGlnIleLeuGlnIleLeuCysValValMet  166

Qy   1105  CGTGAGGGCCTTGAAAATGATATGCTGGCTGTGTTAACCAACAAGCCAATCTGATCCCAGGG  1164
           |||      :::|:|:::||||||||||
Db    167  LeuGlu----------AlaIleGlu  168

Qy   1165  TTGAACCTCAGCGCACTGGCATCTTTTCAACAGGACTGTCCGTGCTATCTCCACCAGCA  1224
           :::              ||||||||
Db    169  ----------------------SerProProLys  172

Qy   1225  GGGCCCCGCCGGAGCTCCCCGTGCCCCTACCACCCC------  1263
           :::||||
Db    173  Gly----------AlaThrIleProTyrArgProLysProAlaSerThrProVal  187

Qy   1264  ----------TTCACTACCCACTCCGGATACTTCTCCAGCCTGTACCCC  1302
           :::|||
Db    188  IlePheAlaLeuGlyLyGlyGlnAlaTyrThrIleGlnGlyVal------AlaIlePro  204

Qy   1303  CAT----------CACCAGTTTGGC----CCGTTCCCGCAT  1329
           |||||       ||||||
Db    205  HisProAspGlnLeuThrLysGlnLeuHisGlnLeuAlaMetGlnGlnThrProPheProPro  224

Qy   1330  -----------------------------CATCACTCTTATCCA  1344
           :::|||||
Db    225  LeuGlyGlnThrAsnProAlaPheProGluLysProLeuHisSerSerGluGlu  244

Qy   1345  GAGCAGGAGATTGTG----------------GAAGGCCCA  1359
           |||||     :::
Db    245  AlaGlnAsnLeuAsnMetGlyGlnSerSerGlyLeuAspAlaSerProAlaSerThrHis  264

Qy   1360  AATCTCTCATCCCAACCCAGGCGTGTGGGCGCCATCATCGGAAGAAGAGGGGGACACATC  1419
           ::|||||||        :::||||||
Db    265  GluLeuThrIleProAsnAspLeuIleGlyCysIleIleGlyArgGlnGlyThrLysIle  284

Qy   1420  AAACAGCTGCGAGATTCGCCGAGCTCTATCAAGATTGCCCCTGG--GAAGGCCCA  1476
           :::||     ||||||||||||
Db    285  AsnGlnIleArgGlnValSerGlyAlaGlnIleLysIleAlaAsnAlaThrGluGlySer  304

Qy   1477  GACGTCAGGCGAAAGGATGGTCATCATCACCGGGACCAGCGAAGCCCAGTTCAAGGCCCAG  1536
           ::||||      ||||:  :::||
Db    305  LysGlyGluThrValLysMetArgGluGluSerGlyAlaArgIleAsnIle------  49
```

```
Qy   1537  GGACGGATCTTTGGGAAACTGAAAGAGGAA  1566
           :::                    ::|
Db    323  TyrLeuIleAsnAlaArgLeuThrSerGlu  332
```

RESULT 6
PCB3_MOUSE

ID   PCB3_MOUSE   STANDARD;   PRT;   339 AA.
AC   P57722;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Poly(rC)-binding protein 3 (Alpha-CP3).
GN   PCBP3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20396135; PubMed=10936052;
RA   Makeyev A.V., Liebhaber S.A.;
RT   "Identification of two novel mammalian genes establishes a subfamily
RT   of KH-domain RNA-binding proteins.";
RL   Genomics 67:301-316(2000).
CC   -!- FUNCTION: Single-stranded nucleic acid binding protein that binds
CC   preferentially to oligo dC (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- TISSUE SPECIFICITY: Ubiquitous.
CC   -!- SIMILARITY: Contains 3 KH domains.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; AF176327; AAG09238.1; -.
DR   HSSP; Q07244; 1KHM.
DR   MGD; MGI:1890470; Pcbp3.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding;
KW   Repeat.
FT   DOMAIN   13   63   KH 1.
FT   DOMAIN   97   150   KH 2.
FT   DOMAIN   261   313   KH 3.
SQ   SEQUENCE   339 AA; 35958 MW; A6F1C7C176A64F9C CRC64;

Alignment Scores:
```
Pred. No.:                                    Length:           339
Score:                         4.94e-08       Matches:           90
Percent Similarity:            244.50         Conservative:      53
Best Local Similarity:         38.65%         Mismatches:       116
Query Match:                   24.32%         Indels:           111
DB:                            3.99%          Gaps:              14
                               1
```

US-09-270-437D-6 (1-3412) x PCB3_MOUSE (1-339)

```
Qy    643  ATTGATTTCCCGCTGCGGATCCTGGTCCCCAGTTGTTGGTGCCATCATCGGAAAG  702
           ::: |||||||:||:|:|::    :::  ||||||:||||||||||||||||||
Db     12  ValThrLeuThrIleArgLeuLeuMetHisGlyLysGluValGlySerIleIleGlyLys  31

Qy    703  GAGGGCTTGACCATAAAGAACATCACTAAGCAGACCCAGTCCCGGTAGATATCCATAGA  762
           ::: |||||||:||    :::  ::::::  ::|:|||||||:||------  ::|
Db     32  LysGlyGluThrValLysMetArgGluValSerGlyAlaArgIleAsnIle------  49
```

```
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DE   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Far upstream element binding protein 1 (FUSE binding protein 1) (FBP).
GN   FUBP1 OR D3ERTD330B.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Retina;
RX   MEDLINE=22382257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy P.H.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE OF 1-286 FROM N.A. (ISOFORM 2).
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   MEDLINE=22354683; PubMed=12466851;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
CC   -!- FUNCTION: Regulates MYC expression by binding to a single-stranded
CC       far-upstream element (FUSE) upstream of the MYC promoter. May act
CC       both as activator and repressor of transcription (By similarity).
CC   -!- SUBUNIT: Interacts with SIAHBP1/FIR and JTV1 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
```

```
Qy    763  AAAGAGAACTCTGGAGCTGCAGAGAGCCTGTCACCATCCATGCCACCCAGAGGGGACT  822
                  ||:::||||| |||||||||||||||
Db     50  ---SerGluGlyAsnCysProGluArgIleValThrIleThrGlyProThrAspAlaIle  68

Qy    823  TCTGAAGCATGCCGCATGATCTT----------GAAATCATGCAGAAAGAGGCA  867
                  ::::|||            :::|||
Db     69  PheLysAlaPheAlaMetIleAlaTyrLysPheGluAsnSerMetSer  88

Qy    868  GATGAGACCAAACTAGCCGGA----GAGATTCCTCTGAAAAATCTTGGCACACAATGGCTTG  924
                  :::    |||  ||:::|||  |||:::||||||
Db     89  AsnSerProAlaThrSerLysProProValThrLeuArgLeuValValProAlaSerGln  108

Qy    925  GTTGAAGACTGATTGGAAAAGAAGGCAGAAAATTGAACAAAATTGACACGAAACACAGG  984
                  ||:::|||||||||||||||||       ||:::|||||
Db    109  CysGlySerLeuIleGlyLysGlyGlySerLysIleLeuLysGluIleArgGluSerThrGly  128

Qy    985  ACCAAGATACAAATCTCATCTTTGCAGGATTTGAGCATATACACCCGAAAGAACCATC  1044
                  :::     |||||     ||||||
Db    129  AlaGlnValGlnValAlaGly------AspMetLeuProAsnSerThrGluArgAlaVal  146

Qy   1045  ACTGTTGAAGGGCAGCAGTTGAGGCCTGTGCCAGTGCTGAGATAGAGATTATGAAGAAGCTG  1104
                  |||||||  :::||||   :::|||||
Db    147  ThrIleSerGlyThrProAspAlaIlePheGlnCysValLysGlnIleCysValValMet  166

Qy   1105  CGTGAGGGCCTTTGAAAAATGATATGCTGGCTGTTAACCAACAAGCCAATCTGATCCCAGGG  1164
                  |||
Db    167  LeuGlu------------------------------------  168

Qy   1165  TTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGACTGTCCGTGCTATCTCCACCAGCA  1224
                  |||||||||
Db    169  ------------------------SerProProLys  172

Qy   1225  GGGGCCCCGCGGAGTCCCCCGTTGCCCCCTACCACCCC---  1263
                  |||
Db    173  Gly----------AlaThrIleProTyrArgProTyrArgProLysProAlaCysThrProVal  187

Qy   1264  TTCACTACCCACTCCGGATACTTCTCCAGCCTGTACCCC  1302
                  :::::|
Db    188  IlePheAlaGlyGlyGlnAlaTyrThrIleGlnIleGlnTyr----------AlaIlePro  204

Qy   1303  CAT------------------------CCGGTTCCGC-AT  1329
                  |||                        ||||||
Db    205  HisProAspGlnLeuThrLysLeuHisGlnLeuAlaMetGlnLeuGlnThrProPhePro  224

Qy   1330  ------------------------CATCACTCTTATCCA  1344
                                          |||||||
Db    225  LeuGlyGlnThrAsnProAlaPheGlyGluLysLeuProLeuHisSerSerGluGlu  244

Qy   1345  GAGCAGGAGATTGTG----------------------  1359
                  |||     :::::::
Db    245  AlaGlnAsnLeuMetGlyGlnSerSerGlyLeuAspAlaSerProProAlaSerThrHis  264

Qy   1360  AATCTTCTCATCCCAACCCAGGCTGTGGCGCCGCCATCATCGGAAGAAGGGGGCCACACATC  1419
                  ::::::|          |||||:::|||
Db    265  GluLeuThrIleProAsnAspLeuIleGlyCysIleIleGlyArgGlnGlyThrLysIle  284

Qy   1420  AAACAGCTGCGCAGATTCGCCGGACGATTGCCCCTGCG--GAAGGCCCA  1476
                  :::::    |||||   :::::||          |||:::--|||||
Db    285  AsnGluIleArgGlnMetSerGlyAlaGlnIleLysLysIleAlaAsnAlaThrGluGlySer  304

Qy   1477  GACGTCAGCGAAAAGGATGATGGTCATCATCACCGGGCCACCCGAAGCCAGTTCAAGGCCCAG  1536
                  ||||||||       :::  :::|||
Db    305  ------SerGluArgGlnIleLeuThrIleThrGlyThrProAlaAsnIleSerLeuAlaGln  322

Qy   1537  GGACGGATCTTTGGGAAACTGAACAGAGAA  1566
                  |||||||||       :::||
Db    323  TyrLeuIleAsnAlaArgLeuThrSerGlu  332
```

RESULT 7
FUB1_MOUSE
ID   FUB1_MOUSE   STANDARD;   PRT;   651 AA.
AC   Q91WJ8; Q8CQY8;

```
CC       Name=1;
CC         IsoId=Q91WJ8-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q91WJ8-2; Sequence=VSP_008322;
CC         Note=No experimental confirmation available;
CC   -!- PTM: Ubiquitinated. This targets the protein for proteasome-
CC       mediated degradation (By similarity).
CC   -!- SIMILARITY: Contains 4 KH domains.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC
DR   EMBL; BC014763; AAH14763.1; -.
DR   EMBL; AK029458; BAC26457.1; -.
DR   MGD; MGI:1196294; D3Ertd330e.
DR   MGD; MGI:1924642; 9530027K12Rik.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR004087; KH dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 4.
DR   SMART; SM00322; KH; 4.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
KW   Transcription regulation; Trans-acting factor; Nuclear protein;
KW   DNA-binding; Repeat; Ubl conjugation; Alternative splicing.
FT   DOMAIN       96    160       KH 1.
FT   DOMAIN      181    247       KH 2.
FT   DOMAIN      271    335       KH 3.
FT   DOMAIN      372    439       KH 4.
FT   DOMAIN       16     21       POLY-GLY.
FT   DOMAIN      345    392       GLY-RICH.
FT   DOMAIN      446    556       PRO-RICH.
FT   VARSPLIC     67     67       D -> DGSWTNPSSTTHWEGMPSPFKD (in isoform
FT                               2).
FT                               /FTId=VSP_008322.
FT   CONFLICT     93     93       MISSING (IN REF. 2).
SQ   SEQUENCE    651 AA;  68539 MW;  F4EECA62FD9FA0D5 CRC64;

Alignment Scores:
Pred. No.:
Score:                  244.00           Length:        651
Percent Similarity:     32.57%           Matches:       162
Best Local Similarity:  23.04%           Conservative:   67
Query Match:             3.9%            Mismatches:    221
DB:                                      Indels:        255
                                         Gaps:           34

US-09-270-437D-6 (1-3412) x FUB1_MOUSE (1-651)

Qy    404 AACAAGTCAACACAGACACAGAAACCGCCGTTGTCAACGTCACATATGCAACAAGA---- 459
              |||          ||||||:::
Db     78 AsnLysSerPheGlyLeuAlaGlnLeuProMet------HisGlnLnSerArg       94

Qy    460 ----GAAGAAGCAAAAATA---                                        474
              ||||||     |||:::
Db     95 SerValMetThrGluGluTyrLysValProAspGlyMetValGlyPheIleIleGlyArg 114

Qy    475 GCCATGAGAAGCTAAGCGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATC    534
              |||:|||||   ||:|::|
Db    115 GlyGlyGlyGluGlnIleSerSerArgIleGlnGlnGluSer--GlyCysLysIleGlnIleAla 133

Qy    535 CCGGATGAAGAG---                   GTGAGCTCCCCTTCGCCCCCT      567
              |||::||
Db    134 ProAspSerGlyLeuProGluArgCysMetLeuThrGlyThrProGluSerVal      153

Qy    568 CAGCGAGCCCAGGCTGGGACCACTCTTCCCGGGAGCAAGGCCAC--GCCCCTGGG---  621
              |||:|||||||
Db    154 GlnSerAlaLysArgLeuGluAspLeuIleValGluLysGlyLeuIleLysArgGlyProAlaProGlyPhe 173

Qy    622 ---GGCACTTCTCAGGCACCAGCACGAGATTGAATTCCCGCTGCGGATCCTGGTC     669
              :::
```

```
Db    174 HisHisGlyAspGlyProGlyAsnAlaValGln---                  GluIleMetIle  188
              |||:::
Qy    670 CCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAACATCACT         729

Db    189 ProAlaSerLysAlaGlyLeuValIleGlyLysValIleGlyGlyGluThrIle---            204
              |||:::|||      ||
Qy    730 AAGCAGACCCAGTCCCGGTAGATATCCATGA---                  AAAGAGAAC         771

Db    205 LysGlnLeuGlnArgAlaGlyValLysMetValMetIleGlnAspGlyProGlnAsn           224
              |||:::
Qy    772 TCTGGAGCTGCAGAGAAGCCTGTCACCATGCCACCCCAGAGGGGACTTCTGAAGCA           831

Db    225 ThrGly--AlaAspLysProLeuArgIleThrGlyAspProTyrLysValGlnGlnAla         243
              :::|||:::|||:::
Qy    832 TGCCGCATGATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAG         891

Db    244 LysGluMetValLeuGluLeuIleArgAspGlnGlyGlyPheArgGluValArgAsnGlu         263
              |||:::
Qy    892 ATTCCTGAAAATCTTGGCACACAATGGCTTG--                    GTT           927

Db    264 TyrGlySerArgIleGlyIleAspValIleAspValProIleProArgPheAlaVal          283
              :::|||
Qy    928 GGAAGACTGATTGGAAAGAGAGGCAGAAATTGAACATGAAACAGGACC                   987

Db    284 GlyIleValIleGlyArgAsnGlyGluMetIleLysLysLysIleIleGlnAsnAspAlaGlyVal 303
              |||:::|||:::
Qy    988 AAGATAACAATCCATCTTTGCAGATTTGAGCATATACAACCCGGAAAGAACCATCACT        1047

Db    304 ArgIleGlnPheLysProAspAspGlyThrThr---                  ProAspArgIleAlaGln 320
              :::|||
Qy   1048 GTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGAGATAGAGATTATGAAGAAGTGCGT       1107

Db    321 IleThrGlyProProAspArgCysGlnHisAla--  AlaGluIleIle                   335
              :::  |||:::
Qy   1108 GAGGCCTTTGAAAATGATATGCTGGCTGTTAACCAACAGCCAATCTGATCCAGGGTTG        1167

Db    336 ---ThrAspLeuLeuArgSerValGlnAlaGlyAsn------  ProGly---              348
                                                       |||
Qy   1168 AACCTCCAGCCGCACTTGGCATCTTTCAACAGGACTGTCCGTGCATCTCCACCAGCAGGG      1227

Db    349                                        GlyProGlyProGly             353
              |||:::
Qy   1228 CCCCGGGA---                         GCTCCCCCGGCTGCCCCTCTAC       1257

Db    354 GlyArgGlyArgGlyArgGlyIleGlyAsnTrpAsnMetGlyProProGlyGly---          371
              |||:::
Qy   1258 CACCCCTTCACTACCCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGGTTGGC     1317

Db    371                                                                   371
Qy   1318 CCGGTTCCCGCATCATCACTCTTATCCAGAGCAGGAGGATTGTGAAATCTCTTCATCCCAACC   1377

Db    372 ---LeuGlnGluPhe--AsnPheIleValProThr                               381
                 |||||||||:::|||
Qy   1378 CAGGCTGTGGGCGCCATCATCGGGAAGGAGAGGGGCCACATCAACAAACAGTCTGGCAGATTC  1437

Db    382 GlyLysThrGlyLysIleIleGlyLysValIleGlyGluThrIleLeuLysSerIleSerGlnGln 401
              |||||||
Qy   1438 GCCGGAGCCCTTTATCAAGATTGCCCCTGCGAAGGCCCACAGGTCAGCGAAGATGGTC       1497

Db    402 SerGlyAlaArgIleGluLeuGlnArgSerProProAsnAlaAspProAsnMetLys         421
              :::||:|||      |||:::
Qy   1498 ATCATCACC---                                                     1506

Db    422 LeuPheThrIleArgGlyThrProGlnIleAspTyrAlaArgGlnLeuIleGluGlu         441
              ::|||  |||
Qy   1507 GGGCACC---                             GGAAGCCAGTT              1526

Db    442 LysIleGlyGlyProValAsnProLeuGlyProProHisGlyProHisGlyVal          461
              :::|||||||
Qy   1527 ---                    CAAGGCCCCAGGGACGGACGATTCTTTGGAAACTGAA     1559
              |||||||:::
```

```
Db    462  ProGlyProHisGlyProProGlyProProGlyThrProMetGlyProTyrAsn 481
              ||||
Qy   1560  AGAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGGAAGCGCA----------  1607

Db    482  ProAla----------------ProTyrAsnProGlyProGlyProAlaProHisGlyPro 497
                                  |||
Qy   1608  --------TATCAGAGTGCCCTTCCACAGCTGGCCGGGTGATTGGCAAAGTGGCAA  1658
                                  |||                       |||
Db    498  ProAlaProTyrAlaPro--------GlnGlyTrpGlyAsnAlaTyrProHisTrpGln  514
              |||                                             |||
Qy   1659  GACCCTGAACGAACTGCAGAACTTAACCAGTGCAGAAGTCATCGTGCCCTGTGACCAAAC 1718
              :::                                         |||
Db    515  GlnGlnAla----------------------------------------------ProPro 519
              :::
Qy   1719  GCCAGATGAAAATGAGGAGGAAGTGATCGTCAGAAATATCGGGCACTCTTTGCTAGCCAGAC 1778
              ||||:::||
Db    520  AspProAlaLysAlaGlyAlaAsp--------------------------ProAsn 529
              ||||
Qy   1779  TGCACAGCG----------------------CAAGATCAGGAGAAATTGTACAACA 1811
              |||
Db    530  SerAlaAlaTrpAlaAlaTyrTyrAlaHisTyrTyrGlnGlnGlnAlaGlnProProPro 549
              |||||||
Qy   1812  GGTGAAGCAGCAGGAGCAGAGAAATACCCTCAGGGAGTGCGCCTCACAGCGCCAGCAAGTGAGG 1871
              ||||||||                          |||
Db    550  AlaAlaProAlaGlyAla-----ProAlaThrThrGlnThrAsnGlyGlnGlyAspGln 567
              ||||||||                          |||
Qy   1872  CTCCCACAGCCACCAGCAAAACAA---------------------------- 1895
              ||||||||
Db    568  GlnAlaProAlaProAlaGlyGlyGlnValAspTyrThrLysAlaTrpGluGluTyrTyrLys 587
              ||||||||
Qy   1896  CGGATGAATGTAGCCCTTCAACACCTGACAGAGAATGAGACCAAACGCAGCCAGCAGATC 1955
              :::|||                            |||
Db    588  LysMetGlyGlnAlaValProAlaPro--------------AlaGlyAlaProAlaPro 601
              :::|||                            |||
Qy   1956  GGGAGCAAACCAAAGACCATCTGAGGAATGAGAAGTCTGCGGAGGCGGCCAGGGACTCTG 2015
              |||||||                          ::|||
Db    602  GlyGlyGlnProAspTyrSerAlaAlaTrpAlaGluTyr-TyrArgGlnAlaAlaTyr 621
              |||||||                          ::|||
Qy   2016  CCGGACCCCTGAGAACCCCAGGGGCGGCGGAGGAGGGGCGGGAAGTCAGCAGGTTTGCCA 2075
              ||||||
Db    622  TyrAlaGlnThrSerProGln--------------------------GlyMetPro 631
              ||||||
Qy   2076  GAACCACCGAGCCCGCCTCCGCCCCCCCAGGGGCTTCTGCAGGCTTCAGCCATCCACTTC 2135
              :::|||                        |||
Db    632  Gln-His--------------ProProAlaProGlnGlyPhe-AlaAsnHisAlaArgSerHisH 648
              :::|||                        |||
Qy   2136  ACCAT 2140

Db    648  isHis 649
```

```
RESULT 8
FUB1_HUMAN
ID   FUB1_HUMAN    STANDARD;    PRT;    643 AA.
AC   Q96AE4; Q12828;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Far upstream element binding protein 1 (FUSE binding protein 1) (FBP)
DE   (FBP1).
GN   (DNA helicase V) (HDH V).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1), SEQUENCE OF 191-193; 203-206;
RP   272-279; 284-290; 300-314; 321-328; 394-397; 409-411; 430-438 AND
RP   440-443, AND FUNCTION.
RC   TISSUE=Leukemia;
RX   MEDLINE=94170991; PubMed=8125259;
RA   Duncan R., Bazar L., Michelotti G., Tomonaga T., Krutzsch H.,
RA   Avigan M., Levens D.;
RT   "A sequence-specific, single-strand binding protein activates the far
```

Right column (references continued):

```
RT   upstream element of c-myc and defines a new DNA-binding motif.";
RL   Genes Dev. 8:465-480(1994).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 45-64; 133-145; 271-283; 308-321; 379-386; 414-424 AND
RP   430-439.
RX   MEDLINE=21127960; PubMed=11222755;
RA   Vindigni A., Ochem A., Triolo G., Falaschi A.;
RT   "Identification of human DNA helicase V with the far upstream
RT   element-binding protein.";
RL   Nucleic Acids Res. 29:1061-1067(2001).
RN   [4]
RP   PARTIAL SEQUENCE, AND MASS SPECTROSCOPY.
RX   MEDLINE=22166132; PubMed=12176931;
RA   Rappsilber J., Ryder U., Lamond A.I., Mann M.;
RT   "Large-scale proteomic analysis of the human spliceosome.";
RL   Genome Res. 12:1231-1245(2002).
RN   [5]
RP   INTERACTION WITH SIAHBP1.
RX   MEDLINE=20337922; PubMed=10082074;
RA   Liu J., He L., Collins I., Ge H., Libutti D., Li J., Egly J.-M.,
RA   Levens D.;
RT   "The FBP interacting repressor targets TFIIH to inhibit activated
RT   transcription.";
RL   Mol. Cell 5:331-341(2000).
RN   [6]
RP   INTERACTION WITH JTV1, UBIQUITINATION, AND PROTEASOME-MEDIATED
RP   DEGRADATION.
RX   MEDLINE=22716800; PubMed=12819782;
RA   Kim M.J., Park B.-J., Kang Y.-S., Kim H.J., Park J.-H., Kang J.W.,
RA   Lee S.W., Han J.M., Lee H.-W., Kim S.;
RT   "Downregulation of FUSE-binding protein and c-myc by tRNA synthetase
RT   cofactor p38 is required for lung cell differentiation.";
RL   Nat. Genet. 34:330-336(2003).
RN   [7]
RP   STRUCTURE BY NMR OF 278-447 IN COMPLEX WITH SINGLE STRANDED FUSE DNA.
RX   MEDLINE=21864720; PubMed=11875576;
RA   Braddock D.T., Louis J.M., Baber J.L., Levens D., Clore G.M.;
RT   "Structure and dynamics of KH domains from FBP bound to
RT   single-stranded DNA.";
RL   Nature 415:1051-1056(2002).
CC   -!- FUNCTION: Regulates MYC expression by binding to a single-stranded
CC       far-upstream element (FUSE) upstream of the MYC promoter. May act
CC       both as activator and repressor of transcription.";
CC   -!- SUBUNIT: Interacts with SIAHBP1/FIR and JTV1.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q96AE4-1; Sequence=Displayed;
```

```
CC   Name=2;
CC     IsoId=Q96AE4-2; Sequence=VSP_008321;
CC   -!- PTM: Ubiquitinated. This targets the protein for proteasome-
CC       mediated degradation.
CC   -!- SIMILARITY: Contains 4 KH domains.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).

DR   EMBL; U05040; AAA17976.2; -.
DR   EMBL; BC017247; AAH17247.1; -.
DR   PDB; 1J4W; 06-MAR-02.
DR   Genew; HGNC:4004; FUBP1.
DR   MIM; 603444; -.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR004087; KH dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 4.
DR   SMART; SM00322; KH; 4.
DR   PROSITE; PS50084; KH_TYPE_1; 4.

KW   Transcription regulation; Trans-acting factor; Nuclear protein;
KW   DNA-binding; Repeat; Ubl conjugation; Alternative splicing;
KW   3D-structure.

FT   DOMAIN     99    163    KH 1.
FT   DOMAIN    184    250    KH 2.
FT   DOMAIN    274    338    KH 3.
FT   DOMAIN    375    442    KH 4.
FT   DOMAIN     13     26    GLY-RICH.
FT   DOMAIN    348    395    GLY-RICH.
FT   DOMAIN    449    559    PRO-RICH.
FT   VARSPLIC  642    643    GQ -> CRFDPASIELAL (in isoform 2).
FT                           /FTId=VSP_008321.
FT   CONFLICT   96     96    Q -> QS (IN REF. 1).
SQ   SEQUENCE  643 AA;  67473 MW;  086D4EAA0ACF807B CRC64;

Alignment Scores:
Pred. No.:                  7.84e-08          Length:        643
Score:                      242.50            Matches:       145
Percent Similarity:         35.48%            Conservative:   81
Best Local Similarity:      22.76%            Mismatches:    241
Query Match:                3.96%             Indels:        170
DB:                           1               Gaps:           27
```

US-09-270-437D-6 (1-3412) x FUB1_HUMAN (1-643)

```
Db   96  nArgSerVal---MetThrGluTyrLysValProAspGlyMetValGlyPheIleIl  115

Qy  696  CGGAAAGGAGGGCTTGACCATTAAAGAACATCACTAGCAGCAGACCCAGTCCCGGGTAGATAT  755
             |||:::         |||||       :::::::              :::::||||
Db  115  eGlyArgGlyIleGlyIleSerArgIleGlnGlnGluSerArgGlyCysLysIleGlnIl  135

Qy  756  CCATAGAAAAAGAACTCTGGAGCTGCGAGAGAAGCCTGTCACCATCCATGCCACCCCAGA  815
             |           |||||            :::       |||||
Db  135  e---AlaProAspSerGlyGlyLeuProGluArgSerCysMetLeuThrGlyThrProGl  154

Qy  816  GGGGACTTCTGAAGCATGCCCGCCATGATTCTTGAAATCATGCAGAAA---  861
             |||   |||||::::::   ::||||||||       :::|||:::
Db  154  uSerValGlnSerAlaLysArgLeuLeuAspGlnIleValGluLysGlyArgProAlaPr  174

Qy  862  ---GAGGCAGATAGAGAACAAACTAGCCGAAGAGATTCCTCTGAAAATCTTGGC  911
                                                       GluIleMetGl
Db  174  oGlyPheHisHisGlyAspGlyProGlyAsnAlaValGln------GluIleMetGl  191

Qy  912  ACACAATGGCTTGGTTGGAAGACTGATTGGAAAAAGAGAGGCAGAAATTTGAAGAAAATTGA  971
             :::|||||||       |||||||           :::||||||||
Db  191  eProAlaSerLysAlaGlyLeuValIleGlyLysLysGlyGlyGluThrIleLysGlnLeuGl  211

Qy  972  ACATGAAACACGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATATACAACCC  1031
             |||        |||||:::       :::|||||               ---GlyPr
Db  211  nGluArgGlyAlaGlyValLysMetValMet------IleGlnAsp---GlyPr  225

Qy 1032  GGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCTGTGCCAGTGCTGAGATAGAGAT  1091
             |||:::                                         ---GlyAlaAsp---
Db  225  oGlnAsnThr------------------GlyAlaAsp-------  231

Qy 1092  TATGAAGAAGCTGCGTGAGCGCCTTTGAAAATGATATGCTGGCTGTGTTAACCAACAAGCCAA  1151
             |||       |||          IleThrGlyAspProTyrLysValGlnGlnGlnLeuLysGl
Db  232  ---LysProLeuArg------IleThrGlyAspProTyrLysValGlnGlnGlnLeuLysGl  248

Qy 1152  TCTGATCCCAGGGTTGAACCTGCAGCGCACTTGGCATCTTTTCAACAGGACTGTCCGTGCT  1211
             :::  :::                |||||
Db  248  uMetVal---------------LeuGluLeuIleGl  254

Qy 1212  ATCTCCACCAGCAGAGGCCCCGGGAGCTCCCCCCCGCTGCCCCCTACCACCCCTTCACTAC  1271
             |||||||      ||||
Db  254  eArgAspGlnGlyGlyPheArgGluVal  263

Qy 1272  CCACTCCGGATACTTCTCCAGCTGTACCCCGTACCCCGTTCCCGCCATCA  1331
             :::::::::|            ||           ArgAsnGluTyrGly
Db  264  ----------------------------------ArgAsnGluTyrGly  268

Qy 1332  TCACTCTTATCCAGACAGCAGAGATTGTGAATCTTCATCCCAACCCAGCTGTGGGCGC  1391
             :::      ::::::::             |||     :::
Db  269  -SerArgIleGlyGlyAsnGluGlyIleAspValProIleProArgPheAlaValGlyIl  288

Qy 1392  CATCATCGGGAAGAAGAGGGGGCACACATCAAACAGCTGGCGAGATTCGCCGGAGCCTCTAT  1451
             :::||||||:::     |||||::::::           :::::
Db  288  eValIleGlyArgAsnGlyGluMetIleLysLysIleGlnAsnAspAlaGlyValArgIl  308

Qy 1452  CAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAGAGGATGGTCATCATCACCGGGCC  1511
             |||||||            :::                |||
Db  308  -------------ThrThrProGluArgIleAlaGlnIleThrGlyPr  327

Qy 1512  ACCGGAA---GCCCAGTTCAAGGCCCAG  1536
             ||||:::   |||||:::|||
Db  327  oProAspArgCysGlnHisAlaAlaGluIleIleThrAspLeuLeuArgSerValGlnAl  347

Qy 1537  --------------GGACGGATCTTTGGAAA  1554
             ||||||:::
Db  347  aGlyAsnProGlyProGlyLeuProGlyGlyArgGlyArgGlyArgGlyGlnGlyAsnTr  367

Qy 1555  ----CTGAAAGAGGAAAACTTCTTAACCCCAAAGAAGAAGT  1592
                 |||||::||
Db  367  pAsnMetGlyProGlyGlyLeuGlnGluPheAsnPhe  381

Qy 1593  GAAGTGGAAGCGCATATCGAGTGCCCTCTTCCACAGCTGGCCGGTGATTGGCAAAGG  1652
             |||:::          :::||||||
Db  382  ------------------ValProThrGlyLysThrGlyLeuIleLeuGlyLysLysGlyI  394
```

Bottom alignment

```
Qy  336  CCCTCCTACCTGCAGTGGGAGGTGTTGGATGGACTTTGGCTCAATATGGGACAGTGGA  395
             :::||||||           :::::::                  GlyGlyGly
Db   10  ProSerSerGlySerAlaGlyValGlyGlyGly------------------GlyGlyGly  22

Qy  396  GAATGTGGAACAAGTCAACACAGAACAGAGAAACCGCCGTTGTCAACGTCACATATGCAAC  455
             |||||           :::||||          :::             AlaAr
Db   23  GlyGlyGly--GlyValAsnAspAlaPheLysAspAlaLeuGlnArg------AlaAr  39

Qy  456  AAGAGAAGAAGCAAAAATAGCCATGGAGAGCTAAGCCGGGCATCAGTTTGAGAACTACTC  515
             :::         :::                  :::
Db   39  gGlnIleAlaAlaLysIleGlyGlyAspAlaGlyThrSerLeuAsnSerAsnAspTyrGl  59

Qy  516  CTTCAAGATTTCCTACATCCGGATGAAGGAGGTGAGCTCCCCTTCGCCCCCTCAGCGAGC  575
             |||||||           |||::         :::
Db   59  yTyrGlyGlnIleLysArgProLeuGluAspGlyGlnProAspAlaLysLysValAl  79

Qy  576  CCAGCCGTGGGACCACTCTTCCCGGGAGCAAGGCCACGCCACGCCCCCTGGGGGCACTTCTCAGGC  635
             |||:::                  ::        LeuProProMetHisGlnGlnGl
Db   79  aProGlnAsnAspSerPheGlyThrGln--------LeuProProMetHisGlnGlnGl  96

Qy  636  CAGACAGATTGATTTCCGCTGCGGATCCTGCGTCCCCACCCAGTTTGTTGGTGCCATCAT  695
             |||       |||              |||||      |||||
```

```
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding; Repeat.
FT   DOMAIN       13     75       KH 1.
FT   DOMAIN       97    162       KH 2.
FT   DOMAIN      279    343       KH 3.
SQ   SEQUENCE    356 AA;  37497 MW;  6D1A61276CA206D CRC64;

Alignment Scores:
Pred. No.:                               Length:            356
Score:                 238.50           Matches:            89
Percent Similarity:    41.78%           Conservative:       66
Best Local Similarity: 23.99%           Mismatches:        137
Query Match:            3.90%           Indels:             79
DB:                                      Gaps:               13

US-09-270-437D-6 (1-3412) x PCB1_RABIT (1-356)

Qy   619 GGGGGCACTTCTCAGGCCAGACAGATTGATTTCCGCTGCGGATCCTGGTCGCCACCCAG 678
             ||| ||:||::||||:|||:   :||||||:|:           ||||||:||  :||::
Db     4 GlyValThrGluSerGlyLeuLeuAsnValThrIleArgLeuLeuMetHisGlyLys  23

Qy   679 TTTGTTGGTGCCATCATCGGAAAGACGAGGGCTTGACCATAAAGAACATCACTAAGCAGACC 738
             |||||||:::|||||||:||||||||:::||||   :||||||:::|||   :::::||||
Db    24 GluValGlySerIleIleGlyLysGlyGlySerLeuIleGlyLysLysLysValLysArgGluGluSer 43

Qy   739 CAGTCCCGGGTAGATATCCATAGAAAGAGAACTCTGAGCGTGCAGAGAAGCCTGTCACC 798
             :::                                 ||||||
Db    44 GlyAlaArgIleAsnIle------SerGluGlyAsnCysProGluArgIleIleThr  60

Qy   799 ATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCGCCGCATGATTCTTGAAATCATGCAG 858
             ::: |||              ::|| :::      :::::          :::::: |||
Db    61 LeuThrGlyProThrAsnAlaIlePheLysAlaMetIleIeIleAspLysLysLeuGlu  80

Qy   859 AAAGAGGCAGATGAGACCCAAACTA------GCCGAAGAGAGATTCCTCTG 900
             ::::::|||     ::: ::|               ::: :||   |||
Db    81 GluAspIleAsnSerSerMetThrAsnSerThrAlaAlaSerArgProValThrLeu 100

Qy   901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAAGAGGCCAGAAATTTG 960
             :::::::||:::                  ::::::       ||:|||:::  |||
Db   101 ArgLeuValValProAlaThrGlnCysGlySerLeuIleGlyLysLysGlyLysLysIle 120

Qy   961 AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020
             |||||:||          ::: :::         |||||:::||||:::|||
Db   121 LysGluIleLeuArgGluSerThrGlyAlaGlnValGlnValAlaGly------AspMetLeu 138

Qy  1021 ATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACCAGTTGAGGCCTGTGCCAGTGCT 1080
             ::: :::|||       |||:::          :::      ||||||:::||
Db   139 ProAsnSerThrGluArgAlaIleThrValLysGlyThrAlaGlyValProGlnSerValThrGluCys 158

Qy  1081 GAGATAGAGATT------ATGAAGAAGTCGCCTGAGCCTTTGAAAATGATATG 1128
             :::                  |||   :::|||:::||||:::     :::
Db   159 ValLysGlnIleCysLeuValMetLeuGluThrLeuSerGlnSerProGlnGlyArgVal 178

Qy  1129 CTGGCTGTTAACCACACAGCCAATCTGATCCCAGGGGTTGAACCTCAGCGCCACTTGGCATC 1188
             ||||||               |||       ::::::      |||   :::|||||||:::
Db   179 MetThrIleProTyrGlnPro------MetProAlaSerSerProValIle 193

Qy  1189 TTTTCAACAGGACTGTCCGTGCTATCTCCACCAGCAGGCCCCGGGAGCTCCCCCGCT 1248
             :::::|||:::              |||:::       ||||||:::
Db   194 CysAlaGlyGlyGlnAspArgCysSerAspAlaAlaGly------ 206

Qy  1249 GCCCCCTACCACCCCTTCACTACCAC------TCCGGTACTTTC 1287
             ||||||   :::|||               :::|||
Db   207 ------TyrProHisAlaThrHisAspLeuGluGlyProLeuAspAlaTyr--- 222

Qy  1288 TCCCAGCCTGTACCCCCATCACCACCAGTTTGGCCCG------ 1320
             :::|||   :::      ||||||:::|||||||:::
Db   223 --SerIleGlyGlnGlyHisThrIleSerProLeuAspLeuAlaLysLeuAsnGlnVal 241

Qy  1321 TTCCCGCCATCATCAC--- 1335
             :::|||||||
Db   242 AlaArgGlnGlnSerHisPheAlaMetMetHisGlyGlyThrGlyThrPheAlaGlyIleIleAsp 261

Qy  1336 --TCTTATCCAGAG--- 1356
```

```
Qy  1653 TGGCAAGACCGTGACGAACTGCAGAAACTTAAACAGTGCAGAGAGTCATCGTGCCTCGTGA 1712
             |||::|::::|||:||:     :::     :||:||          |||   |||:||:||
Db   394 yGlyGluThrIleLysSerIleSerGlnIleSerGlnAlaArgIleGluGlnLeuGlnArgAs 414

Qy  1713 CCAAACGCCAGATGAGAAAATGAGGAAGTG---ATCGTCAGAATTATCGGGCAC--------- 1761
             ::|||:::|||   :::      :::     :::      :::|||||:||
Db   414 nProProAsnAlaAspProAsnMetLysPheThrIleArgGlyThrProGlnGl 434

Qy  1762 ---TTCTTTGCTAGCCAGACTGCCACAGCGCAGGATCAGGGAAATTGTACAACAGGTGAA 1817
                |||       ::::||              :::   ||||||     :::
Db   434 nIleAspTyrAlaArgGlnLeuIleGluGlyLuLysIleGlyGlyProValAsnProLeuGl 454

Qy  1818 GCAGCAGGAGGCAGAAATACCTCAGGGAGTCGGCCTCACAGCGCAGCAAGTGAGGCTCCCA 1877
             |||               :::|||               :::|||
Db   454 yProProValProHisGlyHisGlyVal-ProGlyProHisGlyProProGlyProP 474

Qy  1878 CAGGCACCAGCCAAAACAACGGATGAATGTAGCCCTTCCCACACACCTGACACAGAATGAGACCA 1937
             |||                         |||                :::|||
Db   474 roGly-ProGly-------ThrPro------MetGlyPro 482

Qy  1938 AACGGCAGCCCAGATCGGGAGCAAACCAAAGACCATCTGAGGAATGAGAAGTCTCGCGG 1997
             :::
Db   483 TyrAsnProAlaProTyrAsnProGlyPro--- 492

Qy  1998 AGGGCGGCCAGGGACTCTGCCGAGG---CCCTGAGAACCCCAGGGGCCGAGGAGGCGGG 2054
             |||     ||||||               :::|||   |||       :::|||
Db   493 ------ProGlyProAlaProHisGlyProProAlaProTyrArgAlaProGlnGlyTrpGly 510

Qy  2055 GAAGGTCAGCCAGGTTGCCAGAA------CCACCGAGCCCCGCC 2093
             |||:::   |||||||               :::|||||||
Db   511 AsnAlaTyrProHisTrpGlnGlnGlnAlaProProAspProAla 525
```

```
RESULT 9
PCB1_RABIT

ID   PCB1_RABIT     STANDARD;      PRT;   356 AA.
AC   O19048;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Poly(rC)-binding protein 1 (Alpha-CP1) (hnRNP-E1).
GN   PCBP1.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99207122; PubMed=10101190;
RA   Thiele B.J., Berger M., Huth A., Reimann I., Schwarz K., Thiele H.;
RT   "Tissue-specific translational regulation of alternative rabbit
RT   15-lipoxygenase mRNAs differing in their 3'-untranslated regions.";
RL   Nucleic Acids Res. 27:1828-1836(1999).
CC   -!- FUNCTION: Single-stranded nucleic acid binding protein that binds
CC       preferentially to oligo dC (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: Contains 3 KH domains.
```

```
DR   EMBL; AJ003023; CAA05814.1; -.
DR   HSSP; Q072244; 1KHM.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
```

```
Db   262  SerSerSerProGluValLysGlyTyrTrpAlaSerLeuAspAlaSerThrGlnThrThr  281
          |||          ||||||
Qy  1357  GTGAATCCTTCATCCAACCCAGGCTGTGGGCCCCATCATCGGGAAGAAGAGGGGCACAC  1416
          |||      ::|||||         ::::::||||||
Db   282  HisGluLeuThrIleProAsnAsnLeuIleGlyCysIleIleGlyArgGlnGlyAlaAsn  301
          ||| :||||
Qy  1417  ATCAAACAGCTGGCGAGATTCGCCGGAGCCTCTATCAAGATTGCC---CCTGCGGAAGGC  1473
             : :||||                     ::|||||    |||
Db   302  IleAsnGluIleArgGlnMetSerGlyAlaGlnIleLysIleLeuAlaAsnProValGluIly  321
Qy  1474  CCAGAGCGTCAGCGAAGAGATGGTCATCATCACCCGGCCACCGGAAGCCCAGTTCAAGGCC  1533
             |||                                      ::
Db   322  Ser------SerGlyArgGlnValThrIleThrGlySerAlaAlaSerIleSerLeuAla  339
             |||      :||||
Qy  1534  CAGGGACGACGGATCTTTGGGAAACTGAAAGAGGAA  1566
             ||| :::||||
Db   340  GlnTyrLeuIleAsnAlaArgLeuSerSerGlu  350

RESULT 10
Y475_ARATH

ID   Y475_ARATH     STANDARD;      PRT;   606 AA.
AC   P58223; O49507; Q8H0Y7;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Putative nucleic acid binding protein At4g18375.
GN   At4g18375 OR F28J12.2.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083488; PubMed=10617198;
RA   Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA   Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA   Harris B., Ansorge W., Brandt P., Grivell L., Rieger M.,
RA   Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA   Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
RA   Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA   Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA   Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA   Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA   Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA   Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA   Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA   Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA   Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA   De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA   Van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen S.,
RA   Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA   Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA   Borkova D., Bloecker H., Scharfe M., Grimm M., Loehnert T.-H.,
RA   Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA   Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
RA   Massenet O., Quigley F., Clabauld G., Muendlein A., Felber R.,
RA   Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA   Chefdor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA   Gibbons T., Weber N., Vandenbol M., Bargues M., Terol J., Torres A.,
RA   Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA   Heijnen L., Schwarz S., Scholler P., Heber S., Francs P., Bielke C.,
RA   Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA   Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA   Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA   Sehkon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA   Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA   Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA   Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA   Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W.,
RA   Nelson J., Spieth J., Ryan B., Andrews S., Geisel C., Layman D.,
RA   Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C.,
RA   Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA   Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA   Swaby I.K., O'Shaugnessy A., Rodriguez M., Hoffman J., Till S.,
RA   Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA   Chen E., Marra M.A., Martienssen R., McCombie W.R.;
RT   "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:769-777(1999).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   STRAIN=cv. Columbia;
RX   MEDLINE=22954850; PubMed=14591172;
RA   Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
RA   Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA   Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA   Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA   Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA   Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
RA   Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA   Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
RA   Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
RA   Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA   Satou M., Tamse R., Vayberg M., Wallender E.K., Wong C., Yamamura Y.,
RA   Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT   "Empirical analysis of transcriptional activity in the Arabidopsis
RT   genome.";
RL   Science 302:842-846(2003).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=P58223-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=P58223-2; Sequence=VSP_008899, VSP_008900;
CC         Note=May be due to a competing acceptor site. No experimental
CC         confirmation available;
CC   -!- SIMILARITY: Contains 5 KH domains.
CC   -!- CAUTION: Ref.1 sequences differ from that shown due to erroneous
CC       gene model prediction. At4g18370 and At4g18375 were originally
CC       fused into a single gene.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AL021710; CAA16717.1; ALT_SEQ.
DR   EMBL; AL161548; CAB78839.1; ALT_SEQ.
DR   EMBL; AY133701; AAM91635.1; -.
DR   EMBL; BT001108; AAN64172.1; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_Type_1.
DR   Pfam; PF00013; KH; 5.
DR   SMART; SM00322; KH; 5.
DR   PROSITE; PS50084; KH_TYPE_1; 5.
KW   Hypothetical protein; Nuclear protein; RNA-binding; Repeat;
KW   Alternative splicing.
FT   DOMAIN        35     99       KH 1.
FT   DOMAIN       138    210       KH 2.
FT   DOMAIN       311    380       KH 3.
FT   DOMAIN       394    455       KH 4.
FT   DOMAIN       535    599       KH 5.
FT   VARSPLIC     532    532       L -> F (in isoform 2).
FT                                 /FTId=VSP_008899.
FT   VARSPLIC     533    606       Missing (in isoform 2).
FT                                 /FTId=VSP_008900.
SQ   SEQUENCE   606 AA;  65760 MW;  61F135BBB8647C0C CRC64;
```

Alignment Scores:

Pred. No.:	1.39e-07
Score:	238.50
Percent Similarity:	38.40%
Best Local Similarity:	20.47%
Query Match:	3.90%
DB:	

Length:	606	
Matches:	105	
Conservative:	92	
Mismatches:	216	
Indels:	100	
Gaps:	16	

US-09-270-437D-6 (1-3412) x Y475_ARATH (1-606)

```
Qy   577 CAGGCTGGGGACCACTCTTCCCGGGAGCAAGGCCACGCCCCTGGGGCCACTTCTCAGGCC    636
             :::::|||                    ::::::::
Db     3 GluArgLysLysArgLysGlnIleGlnArgAsnAsnSerGluSerAsnArgGlnIleLys   22

Qy   637 AGACAGAGATTGATTCCCG------------------------CTGCGGATCCTGGTC    669
         |:|::||                                      ||||||
Db    23 ArgArgIleSerHisAspLysAspGluLeuValValTyrArgIleLeuCys          42

Qy   670 CCCACCCAGTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGACATCACT    729
            ||||||||  :::||||||                   ||||||||
Db    43 ProIleAspValValGlyLysValIleGlyLysGlyLysValIleGluAsnAlaIleArg   62

Qy   730 AAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGACTGCAGAGAAG    789
         ||::::::::::    :::                    :::
Db    63 HisAsnThrLysAlaLysIleLysValPhe---AspIleLysHisGlyCysSerGlnArg   81

Qy   790 CCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAGCACATGCCGCATGATTCTTGAA    849
             ||||||::||      :::   :::||                      ::||||
Db    82 ValIleThrIleTyrCysSerValLysSerValLysLysGlyGlnGlyGlu---IleGly   97

Qy   850 ATCATCGAGAAGAGGCCAGATGAGACCAAACTAGCCGAGAGATTCCTCGAAAATCTTG    909
                :::  :::                    ||:::||||||||||:::
Db    98 PheThrLysSerGluAsnGluProLeuCysCysAlaGlnAspAlaLeuLeuLeuLysVal--   116

Qy   910 GCACACAATGGCTTCGTT------------------------------------    927
            ::::::  :::|||
Db   117 --TyrAspAlaIleValAlaSerAspArgGluGluAsnAsnThrLysThrAsnValAspArg   135

Qy   928 ---------------------------------------GGAAGACTGATTGGA    942
                                                 :::
Db   136 AspAspAsnLysGluCysArgLeuLeuValProPheSerGlnSerSerLeuIleGly   155

Qy   943 AAAGAGGCAGAAAATTGAGAAAATTGAACATGAACATGAACAGGACCAAGATAACACAATCTCA    1002
         |||           ||:::||||::||||            :::  ::::
Db   156 LysAlaAlaGlyGluAsnIleLeuLysIleLeuLysArgIleArgArgThrArgAlaSerValLysValVal   175

Qy  1003 TCTTTGCAGATTTGAGCATTATACAACCCGGAAAGAACCATCACTGTGAAAGGGCACAGTT    1062
         :::||||:::|||                              ::::::
Db   176 Ser---LysAspValSer-----AspProSerHisValCysAlaMetGluTyrAspAsn   192

Qy  1063 GAGGCCCTGTCCAGTGCTGAGATAGAGATTATGAAGAAGCTGCCTGAGGCCTTTGAAAAT    1122
         |||||||:::|||::||||:::
Db   193 ValValValIleSerGlyGluProGluSerValLysGlnAlaLeuPheAlaValSerAla   212

Qy  1123 GATATGCTGGCTGTTAACCAACAAGCCAATCTG------------------ATC    1158
            :::||| ::|||  :::|||:::
Db   213 IleMetTyrArgLysIleAsnProArgGluAsnIleProLeuAspSerGlnAspVal   232

Qy  1159 CCAGGGTTGAACCTC------------------------------AGCGCACTTGGC    1185
         ||:::                                       :::||||||
Db   233 ProAlaAlaSerValIleValProSerAspLeuSerAsnSerValTyrProGlnThrGly   252

Qy  1186 ATCTTTTCAACAGGACTGCCGTGCTATCTCCGTGTCGTCCACCAGCAGGGCCCGCGGAGCTCCCCCC    1245
            :::|||                                     ||||||||
Db   253 PheTyrSerAsnGlnAspHisIleLeuLeuGlnGlnGlyAlaGlyValProSerTyrPheAsn   272

Qy  1246 GCTGCCCCCTACCACCCCTTCACTACCCCACTCCGGA------------------    1281
         |||   ||                          :::::::
Db   273 AlaLeuSerValSerAspPheGlnGlyTyrAlaGluThrAlaAlaAsnProValProVal   292

Qy  1282 TACTTCTCCAGCCTGTACCCCCATCACCAGTTTGCCCCGTTCCCGATCATCACTCTTAT    1341
         :::                                        |||||
```

```
Qy   293 PheAlaSerSerLeuProValThrHisGlyPheGly--------GlySerSerArg    308
Db  1342 CCAGACAGGACGGAGATTGTGAATCTTCATCCCAACCCAGGCTGTGGGCGCCATCATCGGG   1401
                                                         :::  :::||||
Qy   309 SerGluGluValPheLysValPheLeuCysProLeuCysAsnIleMetArgValIleGly    328
Db  1402 AAGAGGGGGCCACACATCAAACAGTCGGCGAGATTCGCCGGAGCCTCTATCAAGATTGCC   1461
         |||  ||:::|||:::||||||||||:::                  :::||||:::|||
Qy   329 LysGlyGlySerThrIleLysArgIleArgArgGluAlaSerIleGluGlyValAsn    348
Db  1462 CCTGCCGAA--------GGCCAGAGCAGTCAGCGAAGGATGGTCATCATCACCGGGCCA   1512
             :::||                  :::
Qy   349 AspSerArgThrLysCysSerGlyAspAsp------GluCysValIleIleValThrAlaThr    366
Db  1513 CCGGAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGAAACTGAAAGAGAGGAAAACTTC   1572
         |||                                         :::  :::
Qy   367 GluSerProAspMetLysSerMetAlaValAlaValGluAlaValLeuLeuGlnGluTyr    386
Db  1573 TTTAACCCCAAAGAAGAAGAAGTGAAGCTGGAAGGGCATATCAGATGCTGCCTCTTCCACAGCT   1632
             :::::::||||||                        :::  :::||||
Qy   387 IleAlaAsnAspGluAspAlaLeuAsnValLysMetGlnLeuLeuLeuValSerSerLysValIle    406
Db  1633 GGGCGGGTGATTGGCAAAGGTGGCAAGCACCGTGAACGAACCGACTTAACCAGTCA   1692
         |||           :::||||                     ::||||||||
Qy   407 GlyCysValIleGlyLysGlyLysSerGlyValIleGluAsnGluIleArgLysArgThrAsnAla    426
Db  1693 GAAGTCATCGTGCCTGTGACCAAACGCCAGATGAAAATGAGGAAGTGATCGTCAGAATT   1752
         |||                                         :::  ::||||
Qy   427 AsnIleCysIleSerLysLysGlyLysGlyGluGlu--------LeuValGluVal    441
Db  1753 ATCGGGCACTTCTTTGTTGCCAGACTGCACAGCCGCAGATCAGGAGAATTGTACACAG   1812
                                                     ::||||||||
Qy   442 SerGlyGluVal----------------------------------SerSerValArgAspAlaLeuIleGln    454
Db  1813 GTGAAGCAGCAGGCAGAAAATACCCTCAGGAGTGTCGCCTCACAGCCAGCAAGTGAGGC   1872
         :::  :::                                    ||||||:::||
Qy   455 IleValLeuValArgLeuArgGluAspValIleLeuGlyLysAspValAlaSerValAlaThrArgGly    474
Db  1873 TCCCACAGGCACCCAGCCAAAACAACGATGAATGTAGC   1909
         ||||||                                :::
Qy   474 sProProAla---------ArgThrAspAsnCysSer    483
```

RESULT 11
NOA2_HUMAN
ID   NOA2_HUMAN   STANDARD;   PRT;   492 AA.
AC   Q9UNW9; O43267; Q9UEA1;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   RNA-binding protein Nova-2 (Neuro-oncological ventral antigen 2)
DE   (Astrocytic NOVA1-like RNA-binding protein).
GN   NOVA2 OR ANOVA OR NOVA3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=20197319; PubMed=10735272;
RA   Ueki K., Ramaswamy S., Billings S.J., Mohrenweiser H.W., Louis D.N.;
RT   "ANOVA, a putative astrocytic RNA binding protein gene that maps to
RT   chromosome 19q13.3.";
RL   Neurogenetics 1:31-36(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99007301; PubMed=9789075;
RA   Yang Y.Y., Yin G.L., Darnell R.B.;
RT   "The neuronal RNA-binding protein Nova-2 is implicated as the
RT   autoantigen targeted in POMA patients with dementia.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:13254-13259(1998).
RN   [3]
RP   SEQUENCE OF 29-492 FROM N.A.

```
RA   Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA   Burkhart-Schultz K.J., Gordon L., Dias J., Ramirez M., Stilwagen S.,
RA   Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA   Danganan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,
RA   Attix C., Andreise T., Trankheim M., Amico-Keller G., Coefield J.,
RA   Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,
RA   Arellano A., Brower A., Sanders C., Ow D., Nolan M., Trong S.,
RA   Kobayashi A., Olsen A.S., Carrano A.V.;
RT   "Sequence analysis of a 1.9 Mb region in 19q13.2 between APOE and
RT   D19S412.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 405-480.
RX   MEDLINE=99148126; PubMed=10368286;
RA   Lewis H.A., Chen H., Edo C., Buckanovich R.J., Yang Y.Y.-L.,
RA   Musunuru K., Zhong R., Darnell R.B., Burley S.K.;
RT   "Crystal structures of Nova-1 and Nova-2 K-homology RNA-binding
RT   domains.";
RL   Structure 7:191-203(1999).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 405-491 IN COMPLEX WITH RNA.
RX   MEDLINE=20139984; PubMed=10676814;
RA   Lewis H.A., Musunuru K., Jensen K.B., Edo C., Chen H., Darnell R.B.,
RA   Burley S.K.;
RT   "Sequence-specific RNA binding by a Nova KH domain: implications for
RT   paraneoplastic disease and the fragile X syndrome.";
RL   Cell 100:323-332(2000).
CC   -!- FUNCTION: May regulate RNA splicing or metabolism in a specific
CC       subset of developing neurons (By similarity). Binds single strand
CC       RNA.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- TISSUE SPECIFICITY: Brain. Expression restricted to astrocytes.
CC   -!- DOMAIN: The third KH domain (KH3) recognizes specifically 5'-
CC       UCAY-3'.
CC   -!- DISEASE: Autoantigen in the paraneoplastic opsoclonus myoclonus
CC       ataxia (POMA), a paraneoplastic neurological syndrome/disorder
CC       (PNS/D) associated with breast cancer, fallopian cancer, and SCLCa
CC       and characterized primarily by loss of inhibitory control of motor
CC       neurons in the spinal cord and brainstem. Recognized by the IgG
CC       autoantibody ANNA-2 (also called anti-Ri).
CC   -!- SIMILARITY: Contains 3 KH domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U70477; AAB86611; ALT_INIT.
DR   EMBL; AF083898; AAC72355.1; -.
DR   EMBL; AC006540; AAD13116.1; -.
DR   PDB; 1DTJ; 18-FEB-00.
DR   PDB; 1EC6; 07-APR-00.
DR   Genew; HGNC:7887; NOVA2.
DR   MIM; 601991; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Antigen; Nuclear protein; RNA-binding; Repeat; 3D-structure.
FT   DOMAIN       10     26       BIPARTITE NUCLEAR LOCALIZATION SIGNAL
FT                                (POTENTIAL).
FT   DOMAIN       32     99       KH 1.
FT   DOMAIN      130    196       KH 2.
FT   DOMAIN      232    405       ALA-RICH.
FT   DOMAIN      332    393       GLY-RICH.
FT   DOMAIN      406    473       KH 3.
FT   CONFLICT    247    247       A -> R (IN REF. 1).
FT   CONFLICT    265    267       PAA -> TAT (IN REF. 1).
SQ   SEQUENCE   492 AA;  49008 MW;  41B63EAF689256B CRC64;
```

Alignment Scores:
Pred. No.:   1.41e-07
Score: 238.00
Percent Similarity: 35.85%
Best Local Similarity: 24.03%
Query Match: 3.89%
DB:

Length: 492
Matches: 124
Conservative: 61
Mismatches: 216
Indels: 116
Gaps: 16

US-09-270-437D-6 (1-3412) x NOA2_HUMAN (1-492)

```
Qy    559  TCGCCCCCTCAGCGAGCCAGCGTGGGACCACCTCTTCCCGGAGCAAGGCCACGCCCCT  618
                ::||||||||           :::||||||       :::::||
Db     15  ----ThrProProGluValValCysThrLysArgSerAsnThrGlyLeuGluGly---   31

Qy    619  GGGGGCACTTCTCAGGCCAGACAGATTGATTTCCGCGTGCGGATCCTGGTCCCACCCAG  678
                                          :::::::||||||:::|||||
Db     32  -----------------------------------GluTyrPheLeuLysValLeuIleProSerTyr  42

Qy    679  TTTGTTGGTGCCATCATCGGAAAAGAGAGGGCTTGACCATAAAGAACATCACTAAGCAGACC  738
                ||||||||||:::||||||||       ::: |||||||       ::||:||
Db     43  AlaAlaGlySerIleIleGlyLysGlyGlyGlnThrIleValGlnLeuGlnLysGluThr  62

Qy    739  CAGTCCCGGTAGATATCCATAGAAAAGAGAAC------TCTGGAGCTGCAGAGAAGCCT  792
                                  ::  :::  :::::                ||||||
Db     63  GlyAlaThrIleLysLeuSerLysSerLysAspPheTyrProGlyThrThrGluArgVal  82

Qy    793  GTCACCATCCATGCCACCCCAGGAGGGG---------------ACTTCTGAAGCA  831
                :::                       ||||||||||||||           :: |||
Db     83  CysLeuValGlnGlyThrAlaGlyLeuAlaLeuAlaAsnAlaValHisSerPheIleAlaGluGluLys  102

Qy    832  TGCCGCATGATTCTTGAAATCATGCAGAAA------------  861
                 ||               :||||||||||||||||
Db    103  ValArgGluIleProGlnAlaMetThrLysProGluValValAsnIleLeuGlnProGln  122

Qy    862  ------GAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAATCTTGGCA  912
                       |||      |||:::|||        ||||:::::
Db    123  ThrThrMetAsnProAspArgAlaLysGlnAla------------------LysLeuLeuVal  137

Qy    913  CACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAGAGGCAGAAATTGAAGAAAATTGAA  972
                ||          :::|||||||||||           ::|||
Db    138  ProAsnSerThrAlaGlyLeuIleLleGlyLysGlyGlyAlaThrValLysGlyAlaValMet  157

Qy    973  CATGAAACAGGGACCAAGATAACAATCATCTTTGCAGGATTTGAGCATATACAACCCG  1032
                :::::||:||                  ::: :::||||      ||:||:
Db    158  GluGlnSerGlyAlaTrpValGlnLeuSerGlnLysProGlu---GlyIleAsnLeuGln  176

Qy   1033  GAAAGAACCATCACTGTGAAGGCACAGTTGAGGCCTGTGCCAGTGCTGAGATAGAGATT  1092
                       |||      |||:::         ||
Db    177  GluArgValValThrValSerGlyValGluProGluGlnValHisLysAlaValSerAlaIle  196

Qy   1093  ATGAAGAAGCTGCGTGAGCCCTTGAAAATGATATGCTGGCTGTGTTAACCACAAGCCAAT  1152
                 :::::||:::::::||        :::::::
Db    197  ValGlnLysValGlnGluAspProGlnSerSerSerCysLeuAsnIleSerTyrAlaAsn  216

Qy   1153  CTGATC---CCAGGGTTGAACCTCAGCGCACTTGGCATCTTTCAACAGGACTGTCCGTG  1209
                :::            ||          :::
Db    217  ValAlaGlyProValAlaAsnSerAsnProThrGly-----------------------  228

Qy   1210  CTATCTCCACCAGCACAGGGCCCCCCGGCTGCCCCCCTTCACT  1269
                ||||||||||||||       :::||||||||||||||
Db    229  ---SerProTyrAlaSerProAlaAspValLeuProAlaAlaAlaAlaAlaAlaSerAlaAlaIle  247

Qy   1270  ACCCACTCCGGATAC-------------------TTCTCAGCCTG  1296
                ||||||||                         ||||||
Db    248  AlaAlaSerGlyLeuLeuGlyProAlaGlyLeuAlaGlyValGlyAlaAl  267

Qy   1297  TACCCC-----------------------------------  1303
                |||
Db    267  aLeuProAlaPheSerGlyThrAspLeuLeuAlaIleSerThrAlaLeuAsnThrLeuAl  287

Qy   1304  ------------------------------ATCACCAGTTTGGCCCGTTCCCG  1326
                                               :::        ||||
```

CC   similarity). Acts to regulate a novel glycine receptor alpha-2
CC   chain splice variant (alpha-2N) in developing spinal cord.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Contains 3 KH domains.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; AY262017; AAP20872.1; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH_1; 3.
DR   SMART; SM00322; KH; 3.
KW   Antigen; Nuclear protein; RNA-binding; Repeat.
FT   PROSITE; PS50084; KH_TYPE_1; 3.
FT   DOMAIN   26   42   BIPARTITE NUCLEAR LOCALIZATION SIGNAL
FT   (POTENTIAL).
FT   DOMAIN   48   115   KH 1.
FT   DOMAIN   146   212   KH 2.
FT   DOMAIN   248   384   ALA-RICH.
FT   DOMAIN   396   463   KH 3.
FT   NON_TER   474   474
SQ   SEQUENCE   474 AA;   48443 MW; 70D0926B5BECD131 CRC64;

Alignment Scores:
Pred. No.:   474
Score: 2.02e-07   Length: 474
Percent Similarity: 235.50   Matches: 103
Best Local Similarity: 35.43%   Conservative: 60
Query Match: 22.39%   Mismatches: 162
DB: 3.85%   Indels: 135
                               1   Gaps: 14

US-09-270-437D-6 (1-3412) x NOA1_RAT (1-474)

Qy   562 CCCCCTCAGCGAGCCCAGCGTGCGGACCACTCTTCCCGGAGCAAGCGCCACGCCCCTGGG 621
            |||||| ||| || |:::|| ::: ||| :::
Db   21   PrProAspSerArgLysArgProLeuGluAlaProProGluAlaGly---SerThrLys 39

Qy   622 GGCACTTCTCAGCCCAGACAGATTGATTTCCCGCTGGGATCCTGGTCCCACCCCAGTT   681
            ||:::   ::: ||||||:: :::||||:::||:::
Db   40   ArgThrArgAsnThrGlyGluAspGlyGlnTyrPheLeuValLeuIleProSerTyrAla 59

Qy   682 GTTGGTGCCATCATCGGAAAGAGGAGGGCTTGAACCATAAAGAACATCACTAAGCAGACCCAG 741
            |:::||| ||| :::|||||| ::: ||||:::
Db   60   AlaGlySerIleIleGlyLysGlyGlnThrIleValGlnLeuGlnLysGluThrGly 79

Qy   742 TCCCGGGTAGATATCCATAGAAAAGAGAAC------TCTGGAGCTGCAGAGAAGCCTGTC 795
            ::: ||| ||| ::: :::   :::
Db   80   AlaThrIleLysSerLysSerLysAspPheTyrProGlyThrProGluArgValCys 99

Qy   796 ACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGATGATTCTTGAAAATCATG 855
            ::: ||| :::
Db   100 LeuGluGlyLysThrIleGluAlaLeuAsnAlaValHisGlyPheIleAlaGluLysIle 119

Qy   856 CAGAAAGAGGCAGATGAGCCAAATAGCCGAGAGAATCCCTG------------------- 900
            :::::   :::
Db   120 ArgGluMetProGlnAsnValAlaLysThrGluProValSerIleLeuGlnProGlnThr 139

Qy   901 ----------------------------------AAAATCTTGGCACACAATGGCTTCGGTGGA 930
                                              :::
Db   140 ThrValAsnProAspArgIleLysGlnValLysIleIleValProAsnSerThrAlaGly 159

Qy   931 AGACTGATTGGAAAAGAAGGAGGCAGAAAATTGAACAGAAAATTGAACAGGAACAGGAACAAG 990
            :::::   :::
Db   160 LeuIleIleGlyLysGlyGlyGlyAlaThrValLysAlaIleMetGluGlnSerGlyAlaTrp 179

Qy   991 ATAACAATCTCATCTTTGCAGGATTTGAGCATATACACCCGGAAAAGAACCATCACTCTG 1050
            :::   :::   :::||| |||||:::||||

Db   287 aSerTyrGlyTyrAsnThrAsnSerLeuGlyLeuGlyLeuAsnSerAlaAlaAlaAlaSerGl 307
            :::::   :::
Qy   1327 CATCATCACTCTTATCCAGAGCAGGAGAGATTGTGAATCTCATCCCAACCCAGGCTGCTG 1386
            ::: |||
Db   307 yValLeuAlaValAlaAlaG:yAlaAsnProAlaAlaAlaAlaAlaAlaAsnLeuLe 327

Qy   1387 GGCGCCATCATCGGAAGAAGGGGGCACACATCAAACAGCTGCGAGATTCGCCGGAGCC 1446
            :::||
Db   327 uAlaSerTyrAlaGlyGluAlaGlyAlaGlyAlaGlyProAlaGlyAlaAlaProProProPr 347

Qy   1447 T----------CTATCAAGATTGCCCCTGGGGAGGCCCAGACGTCAGCGAAAGG 1491
            |||
Db   347 oProProGlyAlaLeuGlySerPheAlaLeuAlaLeuAlaAlaAlaAlaAsnGlyTyrLeuGl 367

Qy   1492 ATGGTCATCATCACC-------------------GGGCCACCGGAGAGCCCAGTTCAAGGC 1532
            ::: |||||
Db   367 yAlaGlyAlaGlyGlyGlyAlaGlyGlyGlyGlyProLeuValAlaAlaAlaAlaAl 387

Qy   1533 CCAGGGACGGATCTTTGGGAAACTGAAAGGAGAAACTCTTTAACCCCAAAGAAGAGAGT 1592
            ||
Db   387 aAlaGlyAlaAlaGlyGlyGlyPheLeuThrAlaGlyAlaGluLeuLysS-----AlaAlaGlyAlaGluSerAl 405

Qy   1593 GAAGCTGGAGGCGCATATCAGAGTGCCCTCTTCCACAGTGGCCGGGTGATTGCCAAAGG 1652
            || ::: |||||:::
Db   405 aLysGluLeuValGlyGluIleGlyIleGluAlaValProGluAsnLeuValGlyAlaIleGlyLysGly 425

Qy   1653 TGGCAAGACCGTGAACGAACGTGAAGAACTGCAGAACTTAACCAGTGCAGAGTC--ATCGTGCCTCG 1709
            |:::|| |||| :::
Db   425 yGlyLysThrLeuValGlyThrGlnGlnLeuThrGlyAlaArgIleGlnIleLeuSerLysLysGlu 445

Qy   1710 TGACCAAACGCCAGATGAAAATGAGGAAGTGATCGTCAGAATTATCGGGCACTTCTTTGC 1769
            ::: ::: |||
Db   445 sGlyGluPheLeuProGlyThrArgAsnArgArgValThrIleThrGlySerProAlaAl 465

Qy   1770 TAGCCAGACTGCACACGCCAAGATCAGGGAAATTGTACACACAGTGAAGCAGCAGGAGCA 1829
            |:::||| ||||||
Db   465 aThrGlnAlaAlaGlnTyrLeuIle-----------------SerGlnArgVa 477

Qy   1830 GAAATACCCTCAGGGAGTC---GCCTCACAGCGCAGCAGGTGAGGC 1872
            ||| ::: |||
Db   477 lThrTyrGluGlnGlnGlyValArgAlaSerAsnProGlnLysValGly 492

RESULT 12
NOA1_RAT
ID   NOA1_RAT   STANDARD;   PRT;   474 AA.
AC   Q80WA4;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   RNA-binding protein Nova-1 (Neuro-oncological ventral antigen 1)
DE   (Fragment).
GN   NOVA1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BDIX; TISSUE=Cerebellum;
RA   Knudsen A., Monstad S.E., Vedeler C.A.;
RT   "Nova-1, the paraneoplastic Ri antigen, is associated with breast
RT   cancer.";
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   FUNCTION.
RX   MEDLINE=22120219; PubMed=12124753;
RA   Kumar D.V., Nighorn A., St John P.A.;
RT   "Role of Nova-1 in regulating alpha2N, a novel glycine receptor
RT   splice variant, in developing spinal cord neurons.";
RL   J. Neurobiol. 52:156-165(2002).
CC   -!- FUNCTION: Functions to regulate alternative splicing in neurons by
CC   binding pre-mRNA in a sequence-specific manner to activate exon
CC   inclusion. It binds specifically to the sequence UCAUY (By

```
Db    180  ValGlnLeuSerGlnLysProAsp---GlyIleAsnLeuGlnGluArgValValThrVal  198

Qy   1051  AAGGGCACAGTTGAGCCTGCCAGTGCTGAGATTATGAGAGATTATGAGAGAAGCTGCGTGAG 1110
                ||||           ||| ||  ::|||  ::::||| :|||||
Db    199  SerGlyGluProGluGlnAsnArgLysAlaValGluLeuIleLeuGlnLysIleGlnGlu   218

Qy   1110  ----------------------------------------------------------- 1110

Db    219  AspProGlnSerGlySerCysLeuAsnIleSerTyrAlaAsnValThrGlyProValAla  238

Qy   1110  ----------------------------------------------------------- 1110

Db    239  AsnSerAsnProThrGlyPheTyrAlaAsnThrAlaGluValAlaLeuProThrAlaAla  258

Qy   1111  ------------------------------------------------GCCTTGAA----- 1119
                                                            |||||
Db    259  AlaAlaAlaGlyLeuLeuGlyHisAlaAsnLeuAlaGlyValAlaAlaPheProAlaVal  278

Qy   1120  -----------AATGATATGCTGGCTGTTAACCAACAGCCAATCTGATCCCA         1161
                           |||||:|:||::|  :::
Db    279  LeuSerGlyPheThrGlyAlaAsnAspLeuValAlaIleAlaThrSerAlaLeuAsnThrLeuAla 298

Qy   1162  ---GGGTTGAACCTCAGCGCACTTGGCATCTTTCAACAGGACTGTCCGTGCTATCT     1215
                |||||                     |:::||
Db    299  SerTyrGlyTyrAsnLeuAsnThrLeuGlyLeu--------GlyLeuSer----------  312

Qy   1216  CCACCAGCAGGGCCCGCGGAGCT--------------------CCCCCCGCT         1248
                ||||||                                  |||||
Db    313  ---GlnAlaAlaAlaThrGlyAlaLeuAlaLeuAlaAlaAlaAlaAlaSerAlaAsnProAlaAla 331

Qy   1249  GCCCCCTACCACCCCTTCACTACCCCAC---                           1275
                ||| :::
Db    332  AlaAlaAlaAsnLeuLeuAlaThrTyrAlaAlaSerGluAlaSerGlySerThrAla   351

Qy   1276  ----------------------------------------------------TCC     1278
                                                                :::
Db    352  GlyGlyThrAlaGlyThrPheAlaLeuGlySerLeuAlaAlaAlaAlaThrAlaAlaAlaThrAsn 371

Qy   1279  GGATACTTCTTCCAGCTGTACCCCCATCACCAGTTGGCCCGTTCCCGCATCATCACTCT  1338
                |||||||             ::                 ||||
Db    372  GlyTyrPheGlyAlaAlaSerProLeuAlaAlaSerAlaIleLeuGlyThrGluLysSer 391

Qy   1339  TATCCA---GAGCAGGAGATTGTGAATCTTCATCCCAACCCAGGCTGTGGGCGCCATC  1395
                ||||||               :::|||   |||||| ||||||
Db    392  ThrAspGlySerLysAspValValGlyIleAlaValProGluAsnLeuValGlyAlaIle  411

Qy   1396  ATCGGGAAGAAGAGGGGGCACACATCAAAACAGCTGGCGAGATTCGCCGGAGCCTCTATCAAG 1455
                :::|||||||               :::              ||||:||:
Db    412  LeuGlyLysLysGlyGlyGlyLysThrLeuuValGluTyrGlnGluLeuuThrGlnGlyIleAlaArgIleGln 431

Qy   1456  ATTGCCCCTGCGGAAGGC--------CCAGACGTCAGCGAAAGGATGGTCATCATCACC 1506
                ||||||              |||  ||||||||| |||
Db    432  IleSer--LysLysGlyValProPheValProGlyThrArgAsnArgLysValThrIleThr  450

Qy   1507  GGGGCCACCAGTTCAAGCCCAGTTGGGGACGGGATCTTTGGGAAACTGAAAGAGGAA  1566
                ||||||              :::||||               :::||||:||
Db    451  GlyThrProAlaAlaThrGlnAlaIleThrGlnAlaGlnAsnTyrLeuIleThrGlnAsnArgIleThrTyrGlu 470
```

```
RESULT 13
PBP2_YEAST
ID   PBP2_YEAST  STANDARD;      PRT;    413 AA.
AC   P38151;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   PABI-binding protein 2.
GN   PBP2 OR YBR233W OR YBR1531.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=S288C;
RA   Dubois E., el Bakkoury M., Glansdorff N., Messenguy F., Pierard A.,
RA   Scherens B., Vierendeels F.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   PARTIAL CHARACTERIZATION.
RX   MEDLINE=99038243; PubMed=9819425;
RA   Mangus D.A., Amrani N., Jacobson A.;
RT   "Pbp1p, a factor interacting with Saccharomyces cerevisiae poly(A)-
RT   binding protein, regulates polyadenylation.";
RL   Mol. Cell. Biol. 18:7383-7396(1998).
CC   -!- SUBUNIT: Interacts with PAB1.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- SIMILARITY: Contains 3 KH domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z36101; CAA85196.1; -.
DR   PIR; S46109; S46109.
DR   GermOnline; 138776; -.
DR   SGD; S0000437; PBP2.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat.
FT   DOMAIN       66    130       KH 1.
FT   DOMAIN      148    213       KH 2.
FT   DOMAIN      330    394       KH 3.
SQ   SEQUENCE   413 AA;  45782 MW;  638846509BCE1840 CRC64;
```

```
Alignment Scores:
Pred. No.:                              2.09e-07          Length:            413
Score:                                  235.00            Matches:           102
Percent Similarity:                     40.20%            Conservative:       56
Best Local Similarity:                  25.95%            Mismatches:        151
Query Match:                            3.84%             Indels:             84
Db:                                     1                 Gaps:               16
```

```
US-09-270-437D-6 (1-3412) x PBP2_YEAST  (1-413)

Qy    523  ATTTCCTACATCCGGATGAAGAGGTGAGCTCCCCTTCGCCCCTTCGCCCCTTCGCCCCAGCGAGCCCAGCGT 582
                ||||                                         :::::||
Db     10  IleThrThrThrProThrThrValLeuValSerProAsnThrLeuLysArg--LysLys  28

Qy    583  GGGGACCACTCTTCCCGGAGCAA                                     606
                |||:::                  |||||
Db     29  GlyGluAspThrSerGluGluGlnLeuGlnLeuGluAlaGluIleLeuLysArgValAlaAlaLeuLysAsp  48

Qy    607  ------GGCACACGCCCCTGGGGCACTTCTCAGGCCAGACAGATT------GATTCCCG  654
                      ::                                             |||
Db     49  AlaAspSerHisSerArgSerAsnAspHisAspSerProSerAspAsnValProSerAspValHis  68

Qy    655  CTGGCGGATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGGAGGAGGGCTTGACC  714
                ||||:|:|||||                       :::::|||
Db     69  LeuArgMetLeuCysLeuValLysHisAlaSerLeuIleValGlyHisLysGlyAlaThr  88

Qy    715  ATAAAGAACATCACTAAGCAGACCAGTCCCGGTAGATATCCATAGAAAGAGAACTCT  774
                |:|||:   :::|||  :|||:|:|::||||
Db     89  IleSerArgIleLysSerGluThrSerArgAlaArgIleAsnIleSerAsnAsnIleArg--- 107

Qy    775  GGAGCTGCAGAAGCCGTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGC 834
                ||                                               ||||||
Db    108  GlyValProGluArgIleValTyrValArgGlyThrCysAspAspValThrCysAsnThrGlyAlaTyr 127

Qy    835  CGGCATGATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAA----- 888
```

```
RT   "Characterisation of two major cellular poly(rC)-binding human
RT   proteins, each containing three K-homologous (KH) domains.";
RL   Eur. J. Biochem. 230:447-453(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96016208; PubMed=7556077;
RA   Kiledjian M., Wang X., Liebhaber S.A.;
RT   "Identification of two KH domain proteins in the alpha-globin mRNP
RT   stability complex.";
RL   EMBO J. 14:4357-4364(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lymphocytes;
RX   MEDLINE=94203810; PubMed=8152927;
RA   Aasheim H.-C., Loukianova T., Deggerdal A., Smeland E.B.;
RT   "Tissue specific expression and cDNA structure of a human transcript
RT   encoding a nucleic acid binding [oligo(dC)] protein related to the
RT   pre-mRNA binding protein K.";
RL   Nucleic Acids Res. 22:959-964(1994).
CC   -!- FUNCTION: Single-stranded nucleic acid binding protein that binds
CC       preferentially to oligo dC.
CC   -!- SUBCELLULAR LOCATION: Loosely bound in the nucleus. May shuttle
CC       between the nucleus and the cytoplasm.
CC   -!- TISSUE SPECIFICITY: Abundantly expressed in skeletal muscle,
CC       thymus and peripheral blood leucocytes while a lower expression is
CC       observed in prostate, spleen, testis, ovary, small intestine,
CC       heart, liver, adrenal and thyroid glands.
CC   -!- PTM: PHOSPHORYLATED. THE NON-PHOSPHORYLATED FORM(S) EXHIBITED THE
CC       STRONGEST POLY(RC)-BINDING ACTIVITY.
CC   -!- SIMILARITY: Contains 3 KH domains.
CC   -!- CAUTION: There is probably a frameshift error in the nucleotide
CC       sequence of Ref.3.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X78137; CAA55016.1; -.
DR   EMBL; U24223; AAA91317.1; -.
DR   EMBL; Z29505; CAA82631.1; ALT_FRAME.
DR   HSSP; Q07244; 1KHM.
DR   Genew; HGNC:8647; PCBP1.
DR   MIM; 601209; -.
DR   GO; GO:0005737; C:cytoplasm; NAS.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0003723; F:RNA binding; IDA.
DR   GO; GO:0003697; F:single-stranded DNA binding; IDA.
DR   GO; GO:0016071; P:mRNA metabolism; NAS.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding;
KW   Phosphorylation; Repeat.
FT   DOMAIN       13     75       KH 1.
FT   DOMAIN       97    162       KH 2.
FT   DOMAIN      279    343       KH 3.
FT   CONFLICT    205    205       V -> A (IN REF. 2 AND 3).
SQ   SEQUENCE   356 AA;  37526 MW;  DC85477576DC5104 CRC64;
```

```
Alignment Scores:
Pred. No.:                                            Length:          356
Score:                   2.17e-07                     Matches:          88
Percent Similarity:      234.50                       Conservative:     66
Best Local Similarity:   41.51%                       Mismatches:      138
Query Match:             23.72%                        Indels:           79
DB:                      3.83%                         Gaps:             13
                         1
```

```
Db    128 GlyMetIleValArgAlaLeuLeuGluGluHisGlyAsnGluGluAspAsnGlyGluAspIle 147
          |||||||:::     :::      :::|||
Qy    889 GAGAATTCCTTGAAAATCTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAA    948
          |||||||      ::: :::|||
Db    148 GluIleSerIleAsnLeuLeuIleProHisHisLeuMetGlyCysIleIleIleGlyLysArg 167
          |||:::|:::::||||||
Qy    949 GGCAGAAATTGAAGAAAATGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTG   1008
          |||::::|:::|||
Db    168 GlySerArgLeuArgGluIleGluGluArgAlaAlaAlaLysLeuPheAlaSerProAsn    187
          |||    ::: :::|||||
Qy   1009 CAGGATTTGAGCATATACACACCCGAAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCC   1068
          ::: ::: :::|||:::   |||:::   :::|||
Db    188 Gln------LeuLeuLeuSerAsnAspArgIleLeuThrIleAsnGlyValProAspAla    205
          |||
Qy   1069 TGTGCCGTGCTGAGATAGACATTATGAAGAAGCTG---------------------    1104
          |||
Db    206 IleHisIleAlaThrPheTyrIleSerGlnThrLeuLeuAsnPheGlnMetGluSerPro    225
          |||
Qy   1105 ------------------CGTGAGGCCTTTGAA-----------AATGATATGCTGGCT   1134
          |||  :::
Db    226 GlnLysAsnValLysArgSerIleTyrTyrGlnProThrGlnPheAsnSerValLeuIle    245
          :::  ||||:::
Qy   1135 GTTAACCAACAAGCCAATCTGATC---------------------------------    1158
          ||||    :::
Db    246 AspHisSerGlnProAsnThrIlePheHisGlnIleArgAsnHisGlnTyrHisProSerAsp 265
          |||
Qy   1159 ------------CCAGGGTTGAACCTCAGCGCGCACTTGGCATCTTTCAACAGGA   1200
          |||      ||||:::
Db    266 LysLeuLeuSerTyrLysProAsnLysAsnLeuPro--------IleSerSerThrLeu    282
          ||||:::   ::: :::|||
Qy   1201 CTGTCCGTGCTATCTCCACCAGCAGGGCCCCCGGAGCTCCCCCCGCTGCCCCCTACCAC   1260
          |||||||:::  :::
Db    283 LeuSerMetAlaThrProGlnTyrThrThrAlaSerValAlaAsnAlaThrAlaPheGln    302
          |||||||
Qy   1261 CCC----------TTCACTACCACTCCGGATACTCTCCAGCCTGTACCCC   1302
          |||
Db    303 ProAsnPheValIleProAsnValThrValLeuAspGlyProValIleSer--------    319
          |||
Qy   1303 CATCACCAGTTGGCCCGTCCCGCATCATCAC--------TCTTATCCAGAGCAGGAG   1353
          ::: :::   ||||:::
Db    320 ------------ProAlaProGlyAsnHisLeuLeuMetAsnPheValGlnGlnGlu    334
          |||
Qy   1354 ATTGTGAATCTCTTCATCCCAACCCAGGCTGTGGGCGCCATCATCGGGAAGAAGGGGGCA   1413
          :::|||||||||
Db    335 Ile--------PheIleAspGluLysPheValGlyAsnValIleGlyLysAspGlyLys    351
          :::   ||||:::
Qy   1414 CACATCAAACAGCTGCGCGAGATTCCCGGAGCCTCTATCAAGATTGCC---CCTGCCGAA   1470
          |||||||   :::
Db    352 HisIleAsnSerValLysGluSerThrGlyCysSerIleIleIleGluGlnAspProValGlu 371
          :::|||
Qy   1471 GGCCCAGACGTCAGCCAAAGGATAGGTCATCATCACCGGG   1509
          |||    :::   |||
Db    372 GlySer------SerGluArgArgLeuThrIleArgArgGly    382
```

```
RESULT 14
PCB1_HUMAN
ID   PCB1_HUMAN     STANDARD;      PRT;     356 AA.
AC   Q15365; Q13157; Q14975;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Poly(rC)-binding protein 1 (Alpha-CP1) (hnRNP-E1) (Nucleic acid
DE   binding protein SUB2.3).
GN   PCBP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95331278; PubMed=7607214;
RA   Leffers H., Dejgaard K., Celis J.E.;
```

```
Db    322 Ser------SerGlyArgGlnValThrIleThrGlySerAlaAlaSerIleSerLeuAla 339
Qy   1534 CAGGGACCGGATCTTTGGGAAACTGAAAGAGGAA 1566
              ||| :::|||        :::|||
Db    340 GlnTyrLeuIleAsnAlaArgLeuSerSerGlu 350
```

```
RESULT 15
PCB4_HUMAN
ID   PCB4_HUMAN    STANDARD;    PRT;    403 AA.
AC   P57723;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Poly(rC)-binding protein 4 (Alpha-CP4).
GN   PCBP4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20396135; PubMed=10936052;
RA   Makeyev A.V., Liebhaber S.A.;
RT   "Identification of two novel mammalian genes establishes a subfamily
RT   of KH-domain RNA-binding proteins.";
RL   Genomics 67:301-316(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA   Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA   Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA   Watanabe S., Kimura K., Murakami K., Ishii Y., Kawai Y., Saito K.,
RA   Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA   Ninomiya K., Iwayanagi T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Single-stranded nucleic acid binding protein that binds
CC       preferentially to oligo dC (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: Contains 3 KH domains.
```

```
US-09-270-437D-6 (1-3412) x PCB1_HUMAN (1-356)

Qy    619 GGGGCACTTCTCAGGCCAGACAGATTGATTTCCGCTGCGGATCCTGTCCCACCCAG 678
              |||    |||    :::    ||:|||||||  ::|||
Db      4 GlyValThrGlySerGlySerGlyLeuAsnValThrLeuThrIleArgLeuLeuMetHisGlyLys 23

Qy    679 TTTGTTGGTGCCATCATGGGAAAGGAGGGCTTGACCATAAAGAACATCACTAAGGACAGACC 738
              :::|||||||||||::||||||||||||
Db     24 GluValGlySerIleIleGlyLysLysGlyGluSerValLysLysArgGlyGluGluser 43

Qy    739 CAGTCCCGGTAGATATCCATAGAAAAGAACTCGGAGCTGCAGAGAAGCCTGTCACC 798
              |||:::||||
Db     44 GlyAlaArgIleAsnIle------SerGluGlyAsnCysProGluArgIleIleThr 60

Qy    799 ATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATGATTCTTGAAATCATGCAG 858
              :::    :::|||    :::    :::|||
Db     61 LeuThrGlyProThrAsnIleAlaIlePheLysAlaPheAlaMetIleLeuAspLysLysLeuGlu 80

Qy    859 AAAGAGGCAGATGAGACCAAAACTA--------GCCGAAGAGAATTCCTCTG 900
              :::    :::    |||
Db     81 GluAspIleAsnSerSerMetThrAsnSerThrAlaAlaSerArgProProValThrLeu 100

Qy    901 AAAATCTTGGCACACAATGGCTTGTGTTGGAGACTGATTGGAAAAGAAGGCAGAAATTTG 960
              :::::::::::|||       ||||||||||
Db    101 ArgLeuValValProAlaThrGlnCysGlySerLeuIleGlyLysGlyGlyCysLysIle 120

Qy    961 AAGAAAATTGAACATGAAACAGGGACCCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020
              |||       :::::::::    :::::::::
Db    121 LysGluIleArgGluSerThrGlyAlaGlnIleValAlaGly------AspMetLeu 138

Qy   1021 ATATACAACCCGGAAAGAACCATCACTGTGAAGGGCCAGTGGCCTGTGCCAGTGCT 1080
              :::    |||||||       ||||||:::
Db    139 ProAsnSerThrGluArgAlaIleThrIleAlaGlyValProGlnSerValThrGluCys 158

Qy   1081 GAGATAGAGATT----------ATGAGAAGTGCGTGAGGCCTTTGAAAATGATATG 1128
              :::|||        :::    ::: |||:::
Db    159 ValLysGlnIleCysLeuValMetLeuGluThrLeuSerGlnSerProGlnGlyArgVal 178

Qy   1129 CTGGCTGTTAACCAACAGCCAAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATC 1188
              :::    |||       :::-----MetProAlaSerSerProValIle
Db    179 MetThrIleProTyrGlnPro----------MetProAlaSerSerProValIle 193

Qy   1189 TTTTCAACAGGACTTGCCGTGCTATCTCCACCAGCAGGGCCTCCCCGGCGGAGCTCCCCCCGCT 1248
              |||
Db    194 CysAlaGlyGlyValGlnGlnAspArgCysSerAspAlaValGly---------- 206

Qy   1249 GCCCCCTACCACCCCTTCACTACCCAC------------TCCGGATACTTC 1287
              :::||||||
Db    207 --------TyrProHisAlaThrHisAspLeuGluGlyProLeuAspAlaTyr-- 222

Qy   1288 TCCAGCCTGTACCCCATCACCAGTTGGCCCG-------------- 1320
              :::|||:::
Db    223 --SerIleGlnGlyHisThrIleSerProLeuAlaLeuAsnGlnVal 241

Qy   1321 ------------TTCCCGGATCATCAC------------ 1335
              :::||||
Db    242 AlaArgGlnInSerHisPheAlaMetMetHisGlyThrGlyPheAlaGlyIleAsp 261

Qy   1336 ---TCTTATCCAGAG----------CAGGAGATT 1356
              :::    :::|||
Db    262 SerSerSerProGluValLysGlyTyrTrpAlaSerLeuAspAlaSerProThrGlnThrThr 281

Qy   1357 GTGAATCTCTTCATCCCAACCCAGGCTGTGGGCCCCATCATCGGAAGAAGGGGGCACAC 1416
              :::    |||:::::::::::|||
Db    282 HisGluLeuThrIleProProAsnAsnLeuIleGlyCysIleIleGlyArgGlnGlyAlaAsn 301

Qy   1417 ATCAAACAGCTGGCCAGAGATTCGCCCGGAGCCTCATCAAGATTGCC---CCTGCCGGAAGGC 1473
              :::    :::|||       :::|||||||
Db    302 IleAsnGluIleArgGlnMetSerGlyAlaGlnIleLysLysAlaAsnProValGluGly 321

Qy   1474 CCGAGACGTCAGCGAAAGGATGATGGTCATCATCACCGGGCCACCCGGAAGCCCAGTTCAAGGCC 1533
              :::       |||||||||
```

```
CC    --------------------------------------------------------
DR    EMBL; AF176330; AAG09241.1; -.
DR    EMBL; AK023993; BAB14761.1; -.
DR    EMBL; BC003008; AAH03008.1; -.
DR    EMBL; BC004153; AAH04153.1; -.
DR    Genew; HGNC:8652; PCBP4.
DR    GO; GO:0005634; C:nucleus; ISS.
DR    GO; GO:0003677; F:DNA binding; ISS.
DR    GO; GO:0003723; F:RNA binding; NAS.
DR    GO; GO:0016071; P:mRNA metabolism; ISS.
DR    InterPro; IPR004087; KH_dom.
DR    InterPro; IPR004088; KH_type_1.
DR    Pfam; PF00013; KH; 3.
DR    SMART; SM00322; KH; 3.
DR    PROSITE; PS50084; KH_TYPE_1; 3.
KW    Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding;
KW    Repeat.
FT    DOMAIN         17     67       KH 1.
FT    DOMAIN        101    154       KH 2.
FT    DOMAIN        241    293       KH 3.
SQ    SEQUENCE   403 AA;  41481 MW;  3D99F762A9471265 CRC64;

Alignment Scores:
Pred.No.:                                Length:          403
Score:                231.50             Matches:         89
Percent Similarity:   40.40%             Conservative:    54
Best Local Similarity: 25.14%            Mismatches:      106
Query Match:          3.78%              Indels:          105
DB:                   1                  Gaps:            15

US-09-270-437D-6 (1-3412) x PCB4_HUMAN (1-403)

Qy    619 GGGGGCACTTCTCAGGCCAGACAG------ATTGATTTCCCGCTGCGGATCCTGGTCCCC 672
              ||||||          ::          ||||||  :::||||:::
Db      6 GlyGlyLeuGluGluProGluLeuSerIleThrLeuArgMetLeuMetHis 25

Qy    673 ACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAACATCACTAAG 732
              ||| ||||||||||||||||:|        |||||||  ||| ||| :: ||
Db     26 GlyLysGluValGlySerIleIleGlyLysGluThrLysGlyThrValLysArgIleArgGlu 45

Qy    733 CAGACCCAGTCCGGGTAGATATCCATAGAAAGAGAACTCTGGAGCTGCAGAGAAGCCT 792
              :: ||||:::|||:|||              ::: ||
Db     46 GlnSerSerAlaArgIleThrIle--------SerGluGlySerCysProGluArg 61

Qy    793 GTCACCATCCATGCCACCCAGAGAGGGACTTCTGAAGCATGCCCGCATGATTCTTGAAATC 852
              ||||||||||           |||::|||            ::: ||
Db     62 IleThr------ThrIleThrGlySerThrAlaAlaValPheHisAlaAlaValSerMet 78

Qy    853 ATGCAGAAAGAGGCAGATGAGACCAAACTAGCC-------------------- 885
              ::: ||||||          ||||||||||||
Db     79 IleAlaPheLysLeuAspGluCysAlaAlaProAlaAsnGlyGlyAsnValSer 98

Qy    886 --GAAGAGATTCCTCTGAAAATCTTGGCCACACAATGGCTTGGTTGGAAGACTGATTGGA 942
              ::: |||||||||||||||:||:||||||||
Db     99 ArgProProValThrArgLeuValIleProAlaSerGlnCysGlySerLeuIleGly 118

Qy    943 AAAGAAGGCAGAAATTGAAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCA 1002
              ||||| |::|||:||||||          ::|||||  :::|:|||
Db    119 LysAlaGlyThrLysIleLysGlyLeuLysArgGluThrThrGlyAlaGlnValGlnValAla 138

Qy   1003 TCTTTGCAGGATTTGAGCATATACACCCGAAAGAACCATCACTGTGAAGGGCACAGTT 1062
              |||||          ::: ||||||   :||||||
Db    139 Gly------AspLeuLeuProAsnSerThrGluArgValAlaValThrValSerGlyValPro 156

Qy   1063 GAGGGCCTGTGCCAGTTGCTGAGATAGAGATTATGAAGAAGCTGCGTGAGGGCTTTGAAAAT 1122
              ::: ||||| ||
Db    157 AspAlaIle 159

Qy   1123 GATATGCTGGCTGCTTAACCAACAAGCCAATCTGATCCCAGGGGTTGAACCTGGCGCACTT 1182
              :: ||:|| :::||||
Db    160 ---IleLeuCysValArgGlnIleCysAlaValIle------ 170

Qy   1183 GGCATCTTTTCAACAGGACTGTCCCGTGCTATCTCCACCAGGAGGCCCCGCGGAGCTCCC 1242
```

```
Db    171 --------------------LeuGluSerProProLysGly-- 177
                                 |||||||||    |||
Qy   1243 CCCGCTGCCCCCTACCACCCC--------------------TTCACTACCCACTCC 1278
              ||||||||:::                             ::: |||||
Db    178 AlaThrIleProTyrHisProSerLeuGlyThrValLeuLeuSerAlaAsnGln 197

Qy   1279 GGATACTTCTCCAGCCTGTACCCCCATCACCAGTTGGC-------------------- 1317
              |||:::                                       ::: ||
Db    198 GlyPhe----------SerValGlnGlyGlnTyrGlyAlaValThrProAlaGluVal 213

Qy   1318 ---------------------CCGTTCCGCCATCATCACTCTTATCCA 1344
                                  |||::|||            |||
Db    214 ThrLysLeuGlnGlnLeuSerSerHisAlaValProPheAlaThrProSerValPro 233

Qy   1345 ----------------------GAGCAGGAGGATTGTGAATCTTCATCCCAACC 1377
                                   ::: |:||
Db    234 GlyLeuAspProGlyThrGlnThrSerSerGlnGluProPheLeu--------ValProAsn 250

Qy   1378 CAGGCTGTGGGGCCCATCATCGGGAAGAAGAAGGGGCCACACATCAAACAGCTGGCGGAGATTC 1437
              :::||||||||||||        ::: :::: ||
Db    251 AspLeuIleGlyCysValIleGlyArgGlnGlyserLysIleSerGluIleArgGlnGlnMet 270

Qy   1438 GCCCGAGCCTCTATCAAGATTGCC---CCTGCCGAAGGCCCAGCAGCGTCAGCGAAAGGATG 1494
              :::|||||| |||||                |||||||
Db    271 SerGlyAlaHisIleLysIleGlyIleGlyAsnGlnAlaGluGly------AlaGlyGluArgHis 288

Qy   1495 GTCATCATCACCGGCCACCGGAAGCCCAGTTCAAGGCCCAG 1536
              |||  |||||| :::       :::
Db    289 ValThrIleThrGlySerProValSerIleAlaLeuAlaGln 302

Search completed: July 16, 2004, 10:45:29
Job time : 60.5 secs
```

OM nucleic - protein search, using frame_plus_n2p model

```
            GenCore version 5.1.6
     Copyright (c) 1993 - 2004  Compugen Ltd.
```

Run on: July 16, 2004, 10:30:59 ; Search time 163.5 Seconds
                                    (without alignments)
                        13168.784 Million cell updates/sec

Title: US-09-270-437D-6
Perfect score: 6121
Sequence: 1 ggcacgcgaggaggcgagga.........aaccttgaaaatgtttatt 3412

Scoring table: BLOSUM62
               Xgapop 10.0 , Xgapext 0.5
               Ygapop 10.0 , Ygapext 0.5
               Fgapop 6.0 , Fgapext 7.0
               Delop   6.0 , Delext   7.0

Searched: 1017041 seqs, 31551202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US09270437/runat_16072004_113049_13327/app_query.fasta_1.3591
-DB=SPTREMBL_25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09270437_@CGN_1_1_283_@runat_16072004_113049_13327 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL_25:*
         1: sp_archea:*
         2: sp_bacteria:*
         3: sp_fungi:*
         4: sp_human:*
         5: sp_invertebrate:*
         6: sp_mammal:*
         7: sp_mhc:*
         8: sp_organelle:*
         9: sp_phage:*
        10: sp_plant:*
        11: sp_rodent:*
        12: sp_virus:*
        13: sp_vertebrate:*
        14: sp_unclassified:*
        15: sp_rvirus:*
        16: sp_bacteriap:*
        17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	2806.5	45.9	556	4	Q9Y6M1	Q9y6m1 homo sapien

No.	Score	%	Length	DB	ID	Description
2	2358.5	38.5	545	11	Q7TQF9	Q7tqf9 mus musculu
3	2053.5	33.5	593	13	O57526	O57526 xenopus lae
4	2050	33.5	594	13	O73932	O73932 xenopus lae
5	2033	33.2	576	13	O42254	O42254 gallus gall
6	2023.5	33.1	577	11	Q8CGX0	Q8cgx0 rattus norv
7	2021.5	33.0	577	11	O88477	O88477 mus musculu
8	2015.5	32.9	577	4	Q9NZI8	Q9nzi8 homo sapien
9	2014.5	32.9	577	11	Q8BRH1	Q8brh1 mus musculu
10	2014.5	32.9	577	11	Q80US9	Q80us9 mus musculu
11	1976	32.3	582	13	Q9PW80	Q9pw80 brachydanio
12	1962.5	32.1	579	11	Q8C2J9	Q8c2j9 mus musculu
13	1957.5	32.0	579	11	Q9CPN8	Q9cpn8 mus musculu
14	1950.5	31.9	579	4	O00425	O00425 homo sapien
15	873.5	14.3	580	5	Q8IGK4	Q8igk4 drosophila
16	869	14.2	573	5	Q8IR99	Q8ir99 drosophila
17	866.5	14.2	566	5	Q9VZ69	Q9vz69 drosophila
18	493.5	8.1	828	5	Q21605	Q21605 caenorhabdi
19	337	5.5	100	4	Q86VB1	Q86vb1 homo sapien
20	279.5	4.6	568	10	Q9LXF5	Q9lxf5 arabidopsis
21	277.5	4.5	398	3	O74919	O74919 schizosacch
22	269.5	4.4	833	10	Q9FMF0	Q9fmf0 arabidopsis
23	269	4.4	621	10	Q9C553	Q9c553 arabidopsis
24	267.5	4.4	762	10	Q8LNT9	Q8lnt9 oryza sativ
25	265.5	4.3	680	5	P91393	P91393 caenorhabdi
26	263	4.3	641	5	Q9BLA0	Q9bla0 caenorhabdi
27	262	4.3	351	4	Q8N9K6	Q8n9k6 homo sapien
28	260.5	4.3	169	11	Q7TP50	Q7tp50 rattus norv
29	259	4.2	557	5	Q23487	Q23487 caenorhabdi
30	257.5	4.2	313	4	Q96EP6	Q96ep6 homo sapien
31	254.5	4.2	589	5	Q17935	Q17935 caenorhabdi
32	254.5	4.2	611	5	Q17936	Q17936 caenorhabdi
33	254	4.1	836	10	Q84ZW9	Q84zw9 arabidopsis
34	252	4.1	510	5	Q7Z145	Q7z145 caenorhabdi
35	249.5	4.1	869	10	Q84ZX0	Q84zx0 arabidopsis
36	249	4.1	370	11	Q8BBB0	Q8bbb0 mus musculu
37	248.5	4.1	371	11	Q8C544	Q8c544 mus musculu
38	246.5	4.0	644	10	Q8LDV1	Q8ldv1 arabidopsis
39	246.5	4.0	644	10	Q9FNK3	Q9fnk3 arabidopsis
40	245.5	4.0	318	13	Q7ZVK5	Q7zvk5 brachydanio
41	239.5	3.9	640	10	Q9ASX3	Q9asx3 arabidopsis
42	239	3.9	542	10	Q8S7G1	Q8s7g1 oryza sativ
43	238.5	3.9	532	10	Q8H0Y7	Q8h0y7 arabidopsis
44	237.5	3.9	653	13	Q7ZXS1	Q7zxs1 xenopus lae
45	236.5	3.9	403	11	Q8BP63	Q8bp63 mus musculu

ALIGNMENTS

RESULT 1
Q9Y6M1

```
ID   Q9Y6M1     PRELIMINARY;     PRT;   556 AA.
AC   Q9Y6M1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hepatocellular carcinoma autoantigen.
GN   P62.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99207072; PubMed=10190901;
RA   Zhang J.Y., Chan E.K., Peng X.X., Tan E.M.;
RT   "A novel cytoplasmic protein with RNA-binding motifs is an autoantigen
RT   in human hepatocellular carcinoma.";
RL   J. Exp. Med. 189:1101-1110(1999).
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AF057352; AAD31596.1; -.
DR   GO; GO:0005737; C:cytoplasm; TAS.
DR   GO; GO:0003723; F:RNA binding; TAS.
DR   GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
```

```
DR   GO; GO:0009386; P:translational attenuation; TAS.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR0005504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
DR   PROSITE; PS50102; KH TYPE 1; 4.
SQ   SEQUENCE   556 AA;   61842 MW;   1EDEFB100443DDC4 CRC64;

Alignment Scores:
Pred. No.:              3.05e-200          Length:        556
Score:                  2806.50            Matches:       556
Percent Similarity:     92.82%             Conservative:  0
Best Local Similarity:  92.82%             Mismatches:    0
Query Match:            45.85%             Indels:        43
DB:                     4                  Gaps:          1
```

```
US-09-270-437D-6 (1-3412) x Q9Y6M1 (1-556)

Qy   70 ATGATGAACAAGAAGCTTACATCGGGAACCTGAGCCCGCCGCACCGCCGAGACCTCCGG 129
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MetMetAsnLysLysLeuTyrIleGlyAsnLeuSerProAlaValThrAlaAspAspLeuArg 20

Qy  130 CAGCTCTTTGGGACACGAGAAGCTGCCCCTGGCCGGGACAGGTCCTGCTGAAGTCCGGCTAC 189
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 GlnLeuPheGlyAspArgAspArgAlaGlyArgLeuLeuValLeuGlnLeuLeuLysSerGlyTyr 40

Qy  190 GCCTTCGTGGACTACCCCGACCAGAACTGGGCCATCGGCCATCGAGACCCTTCTCCGGGT 249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41 AlaPheValAspTyrProAspGlnAsnTrpAlaIleArgAlaIleGluThrLeuSerGly 60

Qy  250 AAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGG 309
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LysValGluLeuHisGlyLysIleMetGluValAspTyrSerValSerLysLysLeuArg 80

Qy  310 AGCAGGAAAATTCAGATTCGAAACATCCCTCCACCTGCCAGTGGGAGGTGTTGGATGGA 369
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 SerArgLysIleGlnIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspAsply 100

Qy  370 CTTTTGGCTCAATATGGACAGTTGGACAATGTGGAACAACTCAACACAGACAGAAACC 429
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  101 LeuLeuAlaGlnTyrGlyThrValThrValGluAsnValGluGlnThrAsnThrAspThr 120

Qy  430 GCCGTTGTCAACGTCACATATGCAACAGAGAAGAAGAAAAATAGCCATGAGGAAGCTA 489
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 AlaValValAsnValThrTyrAlaThrArgArgArgGluGluLysMetAlaMetGluLysLeu 140

Qy  490 AGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGATGAAGAGGTG 549
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  141 SerGlyHisGlnPheGluAsnTyrSerPheLysIleSerTyrIleProLeuProAspGluVal 160

Qy  550 AGCTCCCCTTCGCCCCTCAGCGAGCCCGTGGGGACCACTCTTCCCGGGAGCAAGGC 609
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  161 SerSerProSerProLeuSerGluProValGlyAspAlaArgGlyIleHisSerSerArgGluGlnGly 180

Qy  610 CACGCCCTGGGGGCACTTCTCAGGGGCCAGACAGATTGATTCCCCGTTGCCGGATCCTGGTC 669
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 HisAlaProGlyValGlyThrSerGlnAlaArgGlnIleAspPheProLeuArgIleLeuVal 200

Qy  670 CCCACCCAGTTTGTTGGTGCCATCATCCGGAAAGAGGGCTTGACCATAAAAGAACATCACT 729
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  201 ProThrGlnPheValAlaIleIleGlyLysGluGlyLeuThrIleLysGluHisLysAsnIleThr 220

Qy  730 AAGCAGACCCAGTCCCGGGTAGAATATCCATAGAAAAGAAACTTCTGAGCTCAGAGAAG 789
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  221 LysGlnThrGlnSerArgValGluTyrProIleHisGlnArgLysLysSerGluSerGlyAlaAlaGluLys 240

Qy  790 CTTGTAGATCCATCCCACCCACCACGCTTCTTAACCCTAATCCTCTCCTATGTTCGTCCGT 849
```

```
Qy  850 ATCATGCAGAAAGAGGCAGATAGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAATCTTG 909
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  261 IleMetGlnLysGluAlaAspGluThrLysLeuAlaGluGluIleProLeuLysIleLeu 280

Qy  910 GCAACAACAATGGCTTGGTTGGTAGGAACTGATTGGAACAAAGAAGGCAGAAATTTGAAGAAATT 969
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  281 AlaHisAsnGlyLeuValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLysLysIle 300

Qy  970 GAACATGAAACACGGACCAAGATAACAATCCATCTTTGCAGGATTTGAGCATATACAAC 1029
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GluHisGluThrGlyThrLysIleThrIleSerSerLeuGlnAspLeuSerIleTyrAsn 320

Qy 1030 CCGGAAAGAAACCATCACTGTGAAGGGCCACAGTTGAAGCCTGCCAGTGCTGAGATAGAG 1089
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  321 ProGluArgThrIleThrValLysGlyThrValGluAlaCysAlaSerAlaGluIleGlu 340

Qy 1090 ATTATGAAGAAGCTGCGTGAGCCCTTTGAAAATGAATATGGCTGTGTTAACCAACAAGCC 1149
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  341 IleMetLysLysLeuArgGluAlaPheGluAsnAspMetLeuAlaValAsn--------- 357

Qy 1150 AATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTCAACAGGACTGTCCGTG 1209
           ------------------------------------------------------------
Db  357 ------------------------------------------------------------ 357

Qy 1210 CTATCTCCACCAGCAGGCCCCCGGCCCGAGCCTCCCCCCGTCCCCTACCACCCCTTCACT 1269
           ------------------------------------------------------------
Db  357 ------------------------------------------------------------ 357

Qy 1270 ACCCACTCCGGATACTTCTCCAGCCCTGTACCCCCATCCAGTTTGGCCCGTTCCCGCAT 1329
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  358 ThrHisSerGlyTyrPheSerSerLeuTyrProHisHisGlnPheGlyProPheProHis 377

Qy 1330 CATCACTCTTATCCAGAGCCAGAGATTGTGAATCCTTCATCCCAACCCAGGCTGTGGGC 1389
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  378 HisHisSerTyrProGluGlnGluIleValAsnLeuPheIleProThrGlnAlaValGly 397

Qy 1390 GCCATCACTCGGGAAGAAGGGGCCACACATCAAACAGCTGCGGAGATTCGCCGGAGCCTCT 1449
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  398 AlaIleIleGluGlyLysLysGlyAlaHisIleLysIleLeuAlaArgPheAlaGlyAlaSer 417

Qy 1450 ATCAAGATTGCCCCTGCGGAAGGCCCAGAGCTCAGCGAAGGATGGTCATCATCACCGGG 1509
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  418 IleLysIleAlaProAlaGluGlyProAspSerGluArgMetValIleIleIleThrGly 437

Qy 1510 CCACCCGAAGCCCAGTTCAAGGCCCAGGGACGATCTTTGGGAAACTGAAAGAGGAAAAC 1569
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  438 ProProGluAlaGlnPheLysAlaGlnGlyArgIlePheGlyLysLeuLysGluGluAsn 457

Qy 1570 TTCTTTAACCCAAAGAAGAAGTGAAGCTGAAGCCCATATCAGAGTGCCCTTCCACA 1629
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  458 PhePheAsnProLysLysLysLeuGluValLysLeuGluAlaHisIleArgValProSerSerThr 477

Qy 1630 GCTGGCCGGGTGATTGGCAAAGTGGCAAAGTCCGCAGAACTGCAGAACTTAACCAGT 1689
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  478 AlaGlyArgValIleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuThrSer 497

Qy 1690 GCAGAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGGATGGTCAGA 1749
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  498 AlaGluValIleValAlaProGlnThrProAspGluAspAspGluAsnGluLysValIleValArg 517

Qy 1750 ATTATCGGGCACTTCTTTGCTAGCCAGACTCCACAGCGCCAAGATCAGGGAAATTGTACAA 1809
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  518 IleIleGlyHisPhePheAlaSerGlnThrAlaGlnArgLysIleArgGluIleValGln 537

Qy 1810 CAGGTGAAGCAGCAGGAGCAGAAAATACCCTCAGGGAGTCCGCCTCACAGCCAGCAAG 1866
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  538 GlnValLysGlnGlnGluGlnLysIleTyrProGlnGlyValAlaSerGlnHisArgSerLys 556

RESULT 2
```

```
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC054552; AAH54552.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE 545 AA; 59679 MW; 88DD35F30DF3091 CRC64;

Alignment Scores:
Pred. No.:
Score:                  2358.50            Length:           545
Percent Similarity:     95.71%             Matches:          456
Best Local Similarity:  93.06%             Conservative:      13
Query Match:            38.53%             Mismatches:        14
DB:                                        Indels:             7
                        7.16e-167          Gaps:               1

US-09-270-437D-6 (1-3412) x Q7TQF9 (1-545)
```

```
Qy    73 ATGAACAAGCTTACATCGGACAACTGGACTCGAGCCCCGCCGTCCACCGGCCGAGCACCTCCGGCAG 132
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MetAsnLysLeuTyrIleGlyTyrIleGlyTyrIleGluGlyLeuAsnLeuSerProAlaValThrAlaAspLeuArgGln 20

Qy   133 CTCTTTGGGGACAGGAAGTGCCCCTGGCGGGACAGGTCCTGCGTCTGAAGTCCGGCTACGCC 192
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    21 LeuPheGlyAspArgLysCysProTrpArgAspArgSerCysValLeuLysSerGlyTyrAla 40

Qy   193 TTCGTGGACTACCCCGACCAGGAACTGGGCCATCCGCGCCATCAGGACCCTCTCGGGTAAA 252
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    41 PheValAspTyrProAspGlnGluLeuGlyHisProArgHisGlnAspProLeuGlyLys 60

Qy   253 GTGGAATTGCATGGGAGAAAATCATGCAAGAAGTTGATTACTCCTCTAAAAAGCTAAGAGAGC 312
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 ValGluLeuHisGlyArgLysSerCysLysLysLeuIleThrProLeuLysSerLysGlu 80

Qy   313 AGGAAAATTCAGATCGAAAACATCCTCTCACCTGCAGTGGGAGGTGTTGGATGGACTT 372
         ||:::||||||:||||||||:||||||||||||||||||||||||||||||||||||
Db    81 ArgArgIleGlnIleGluArgAsnIleProHisProLeuGlnTrpGluValLeuAspGlyLeu 100

Qy   373 TTGGCTCAATATGGACCAGTGGGAGAATGTGGAACAGTCAACAGACACAGAAACCCGCC 432
         ||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   101 LeuAlaGluTyrGlyThrValGluAsnValGluGlnValAsnThrAspThrGluThrAla 120

Qy   433 GTTGTCAACGTCACATATGCAACAAGAGAGAAGAAGCAAAATAGCCATGAGAGAGCTAAGC 492
         |||||||| ||||||||||||:::||||||||||||::::||||||||
Db   121 ValValAsnValThrTyrMetThrArgGluValAlaLysLeuAlaIleGluLysLeuSer 140

Qy   493 GGGCATCAGTTGAGAACTACTCCTCAAGATTCCTACATCCGGATCAAGAAGAGTGAGC 552
         ||||||:::||||:::||||||||||||||||||||||||||||||||||||||||||
Db   141 GlyHisGlnPheGluAspTyrSerPheLysIleSerTyrIleProAspAspGluValSer 160

Qy   553 TCCCCTTGCCCCCTCAGCGAGCCAGCGTGGGACCACTCT----CCCGGGAGCAAGGCCAC 612
         ||||||||||||||||||||||||||||||||||||||                    
Db   161 SerProSerProProHisArgAla--------------------ArgGluGlnGlyHis 173

Qy   613 GCCCCTGGGGCACTTCTCCGGCCAGACAGATGATTTCCGCTGCGGATCCTGGTCCCC 672
         |||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db   174 GlyProGlySerSerGlnAlaArgGlnIleAspPheProLeuArgIleLeuValPro 193

Qy   673 ACCCAGTTGTTGGTGCCATCATCCGGAAGGAGGGGCTGACCATAAAGAACATCACTAAG 732
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   194 ThrGlnPheValGlyAlaIleIleGlyLysGluGlyLeuThrIleLysAsnIleThrLys 213

Qy   733 CAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCGGAGCTGCAGAGAAGCCT 792
         ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db   214 GlnThrGlnSerArgValAspIleHisArgLysGluAsnSerGlyAlaAlaGluLysPro 233

Qy   793 GTCACCATCCATGCGGGACTTCTGAAGCATGCCGCCATGATTCTTGAAATC 852
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   234 ValThrIleHisAlaThrProGluGlyThrSerGluAlaCysArgMetIleLeuGluIle 253

Qy   853 ATGCAGAAAGAGAGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAATCTTGGCA 912
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   254 MetGlnLysGluSerAspGluThrLysLeuAlaGluAlaGluValProLeuLysIleLeuAla 273

Qy   913 CACAATGGCTGGTTGGAAGACTGATGGAAAGAAGGCAGAAATTGAAGAAAATTGAA 972
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   274 HisAsnGlyPheValGlyLeuIleGluGlyLysGluGlyLysArgAsnLeuLysIleGlu 293

Qy   973 CATGAAACAACAGGGACCAAGATAACAATCCATCTTGAAGAATGATATGCTGGCTGTGTTAACCAACAAGCCAAT 1032
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   294 HisGluThrGlyThrLysIleThrIleSerProGluGluSerLeuGlnAspLeuSerIleTyrAsnPro 313

Qy  1033 GAAAGAACCATCACTGTGAAGGCACAGTTGAGGCCTGCCAGTGCTGAGATAGAGATT 1092
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   314 GluArgThrIleThrValLysGlyThrGluArgLeuGluAlaCysLeuAsnAlaGluIleGlu 333

Qy  1093 ATGAAGAAGCTGCGTGAGCCTTGAAAATGATATGCTGGCTGTGTTAACCAACAAGCCAAT 1152
         ||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||
Db   334 MetLysLysLeuArgGluAlaPheGluAsnAspMetLeuAlaValAsnGlnGlnAlaAsn 353

Qy  1153 CTGATCCCAGGGTTGAACCTCAGCGCCACTGGCCATCTTTCAACAGACTGTCCGTGCTA 1212
         ||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db   354 LeuIleProGlyLeuAsnLeuSerAlaLeuGlyIlePheSerThrGlyLeuSerValLeu 373

Qy  1213 TCTTCCACCAGCAGGCAGGAGTTGAATCTCTTCATCCCAACCCCTTCACTACC 1272
         |||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db   374 ProProAlaGlyTyrProArgGlyValProProSerProProTyrHisProPheAlaThr 393

Qy  1273 CACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTGCCCGTTCCCGCATCAT 1332
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   394 HisSerGlyTyrPheSerSerLeuTyrProHisHisHisGlyProPheProHisHis 413

Qy  1333 CACTCTTATCCAGACAGGAGATTGTGAATCTCTTCATCCCAACCAGGCTGTGGGCGGCC 1392
         |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db   414 HisSerTyrProGluGlnValSerLeuPheLeuProThrGlnAlaValGlyAla 433

Qy  1393 ATCATCGGGAAGAAGGGGGCACCACATCAAACAGCTGGCGAGATTCGCCGGAGCCTCATC 1452
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   434 IleIleGlyLysLysGlyAlaHisIleLysIleLysGlnLeuAlaArgPheAlaGlyAlaSerIle 453

Qy  1453 AAGATTGCCCCTGCCGAAGGCCCCAGAGTCAGCGAAAGGATGGTCATCATCACCGGGCCA 1512
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   454 LysIleLeuAlaProAlaGluGlyProAspValSerGluArgMetValIleIleThrGlyProPro 473
```

```
Qy    1513  CCGGAAGCCCAGTTCAAGGGCCCAGGGACGG 1542
            ||||||||||||||||||||||||||||
Db     474  ProGluAlaGlnPheLysValSerAlaLys 483

RESULT 3
O57526
ID   O57526           PRELIMINARY;     PRT;       593 AA.
AC   O57526;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   KH domain-containing transcription factor B3.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92249652; PubMed=1577195;
RA   Pfaff S.L., Taylor W.L.;
RT   "Characterization of a Xenopus oocyte factor that binds to a
RT   developmentally regulated cis-element in the TFIIIA gene.";
RL   Dev. Biol. 151:306-316(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Griffin D., Taylor W.L.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Havin L., Git A., Elisha Z., Oberman F., Yaniv K.,
RA   Pressman Schwartz S., Standart N.M., Yisraeli J.K.;
RL   Genes Dev. 0:0-0(1998).
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AF042353; AAB97457.1; -.
DR   EMBL; AF064633; AAC18597.1; -.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
SQ   SEQUENCE 593 AA;   65385 MW;  5A5AB4B4A1D55DF7 CRC64;

Alignment Scores:
Pred. No.:                          Length:        593
Score:               3.8e-144       Matches:       403
Percent Similarity:  2053.50        Conservative:  75
Best Local Similarity: 79.67%       Mismatches:    87
Query Match:         67.17%         Indels:        35
DB:                  33.55%         Gaps:          5
                     593 AA;    (1-593)
                     13

US-09-270-437D-6 (1-3412) x O57526 (1-593)

Qy     73  ATGGAACAAGCTTTACATCGGGAACCTGAGCCCGGGACTTCCGGCAG 132
            ||||||||||||||||||||||||||||||
Db      1  MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnValSerProProAspLeuGluSer 20

Qy    133  CTCTTTGGGGACAGGAAGCTGCCCGGGGACCAGGTCCTGCTGAAGTCCGGCTAGGCC 192
            |||:  |||   |||||  ||||||||||||||| |||
Db     21  LeuPheLysGluSerGlyIleProPheThrGlyGlnPheLeuValLysSerGlyTyrAla 40

Qy    193  TTCGTGGACTACCCCGACCAACTGGGCCATGCGGCCATCGAGACCCTTCGGGTAAA 252
            ||||   |||:  |||||||||||||||||||||||||||
Db     41  PheValAspCysProGluThrTrpAlaMetLysAlaIleAspThrLeuArgGlyLys 60

Qy    253  GTGGAATTGCATGGAAAATCATGGAAGTGATTACTCAGTCTCTAAAAAGCTAAGGAGC 312
```

```
Db    409 ----------------------------------------------------------------ProProPheGlyGlnProGluSerGlu 418

Qy   1354 ATTGTGAATCTCTTCATCCCAACCCAGGCTGTGGGCGCCATCATCGGGAAGAAGGGGCA 1413
                |||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    419 ThrValHisLeuPheIleProAlaLeuAlaValGlyAlaIleIleGlyLysLysGlyGln 438

Qy   1414 CACATCAAAACAGCTGCGAGATTCGCCGGAGCCCTCATCAAGATTGCCCCTGCGAAGGC 1473
                |||:|||||||||||||||||||||||||||::||||||||||||||||||||||||||
Db    439 HisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaProAlaGluGly 458

Qy   1474 CCAGACGTCAGCGAAGGATGGTCATCATCACCGGGCACCGGAAGCCCAGTTCAAGGCC 1533
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    459 ProAspAlaLysLysArgMetValIleIleThrGlyProProLysLysGlnPheLysAla 478

Qy   1534 CAGGGACGGATCTTTGGGAAACTGAAAGAGAGGAAAACTTCTTAACCCCAAAGAAGAAGTG 1593
                |||||||||||||||||||||:|||||||||||||||||||||::|||||||||||||
Db    479 GlnGlyArgIleTyrGlyLysLeuLysGluGluArgAsnPhePheGlyProLysGluVal 498

Qy   1594 AAGTTGGAAGGCCATATCAGAGTGCCCTCTTCACCAGCTGCCGGGTGATTGGCAAAGGT 1653
                ||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db    499 LysLeuGluGluAlaHisIleLysValProSerTyrAlaAlaAlaGlyValIleGlyLysGly 518

Qy   1654 GGCAAGACCGTGAACGAACTGCAGAACTTAACCAGTGCAGAAGTCATCTGCCTCGTGAC 1713
                |||||||||||||||:||||||||||||||||||:|||||||||||::|||||||||||
Db    519 GlyLysThrValAsnGluLeuGlnAsnLeuThrSerAlaGluValIleCysLeuValProArgAsp 538

Qy   1714 CAAACGCCCAGTGAAAATGAGGAAGTGATCGTCAGAATTATCGGGCCACTTCTTTGCTAGC 1773
                |||:||||||:||||||:|||||||||||||||||:||||||||||:||||||||||||
Db    539 GlnThrProAspGluAsnAspValValIleValIleThrGlyHisPheTyrAlaSer 558

Qy   1774 CAGACTGCACAGCGCAAGATCAGGGAAATTGTACACACAGGTGAAGCAGCAGAGCAGAAA 1833
                ||||:||||||||||||||||:||||||||:::|||||||||||||||||:::|||||
Db    559 GlnLeuAlaGlnArgLysIleArgGluIleGlnGlyLysLeuLeuAlaGlnArgArgGlnGlnGln 578
```

```
Alignment Scores:
Pred. No.:                    7.07e-144          Length:                594
Score:                        2050.00            Matches:               405
Percent Similarity:           79.24%             Conservative:          76
Best Local Similarity:        66.72%             Mismatches:            92
Query Match:                  33.49%             Indels:                34
DB:                           13                 Gaps:                  6

US-09-270-437D-6 (1-3412) x O73932 (1-594)
```

```
Qy     73 ATGAACAAGCTTACATCGGAACCTGACCTGAGCCCCGTCACCCGCAGCGAGACCTCCGGCAG 132
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetAsnLysLeuTyrIleGlyThrAspLeuSerProThrAspSerGluSer 20

Qy    133 CTCTTGGGGACAGGAGAGTGCCCCTGCCGGGGACAGGTCCTGCTGAAGTGCCGGCTACGGC 192
                |||:::||||    |||||||||||||||||||||||||||||||||||||| :::
Db     21 LeuPheLysGluSerLysIleProPheThrGlyGlnPheLeuValLysSerGlyTyrTyrAla 40

Qy    193 TTCGTGGACTACCCCGACCAGAACTGGGCCATGGCCATCGAGACCCTCCGGGTAAA 252
                ||||||||||||:::| ||||||:::|||||| |||:|:|||||||||||||||||
Db     41 PheValAspCysProAspGluThrTrpAlaMetLysAspThrIleuSerGlyLys 60

Qy    253 GTGGAATTGCATGGGAAAATCATGGAAGTGATTACTGATTGCTCTAAAAAGCTAAGGAGC 312
                ||||||||||||||||||:::||||||||||||||||||||||||:::||||||||||
Db     61 ValGluLeuHisGlyLysValIleGluValIleHisSerValProLysArgGlnArgSer 80

Qy    313 AGGAAAATTCAGATTGAAACATCCCTCCACCTGCAGTGGAGGTGTTGGATGGACTT 372
                ||||||||:::||||||||| ||||||:::|||||||||||||:::||||||||||
Db     81 ArgLysLeuGlnIleGluArgAsnIleProProHisLeuGlnTrpGlnValLeuAspSerLeu 100

Qy    373 TTGGCTCAATATGGACAGTGGAGAATGTGGAACAAGTCAACAACAGACAGAAAACCGCC 432
                ||||||||:::| |||||:::|||||||||||||||||||||||||||||:::||||||
Db    101 LeuAlaGlnTyrGlyThrValGluAsnCysGluGlnValAsnThrGluGlnThrAla 120

Qy    433 GTTGTCAACGTCACATATCCAACAAGAGAAGAAGAAGAAGCAAAATAGCCATGGAGACTAAGC 492
                ||||||||||||||||||||:::::||||||:||||||||||||| :::|||||:
Db    121 ValValAsnValThrTyrAlaAsnLysAlaArgGlnGlyLeuGluLysLeuAsn 140

Qy    493 GGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGATCGAAGAGGTGAGC 552
                |||::|||  |||||||||||||||||||:::|||||||:::||||||||||  :::
Db    141 GlyTyrGlnLeuGluAsnTyrSerLeuLysValThrTyrIleProAspGluMetAlaThr 160

Qy    553 TCCCCTTGCGCCCCCTCAGCGAGCCCAGCGTGGGAGACCACTCTTCCCGGGAGCAAGGCCAC 612
                |||||||||||||:::||||||||||||:::||||||:::||||||:::|
Db    161 ProGlnAlaProSerGlnLeuGlnLeuGlnGlnGlnGlnGlnHisProGlnGlyArg 180

Qy    613 ----------------------------------------GCCCCTGGGGCACTTCTCAGGCG 636
                                                         ::||||   :::::
Db    181 ArgGlyPheGlyGlnArgGlyProAlaArgGlnGlySerProGlyAlaAlaArgPro 200

Qy    637 AGA---CAGATTGATTTCCCGCTGCGGATCCCCCACCCAGTTTGTTGGTGCCATC 693
                ::: |     |::: :::|||||||||:::|||||||||||||||||:|||
Db    201 LysProGlnThrGluValProLeuArgMetLeuValProThrHrGlnInPheValAlaIle 220

Qy    694 ATCCGAAAGAGGAGGCCTTGACCATAAAGAACATCACTAAGGACCAGGTCCCGGGTAGAT 753
                |||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db    221 IleGlyLysGlyIleAlaThrIleAlaGluAsnIleThrLysGlnInSerLysIleAsp 240

Qy    754 ATCCATAGAAAAGAGAACTCTGGAGCTGCAGAGAACCTGTCACCATGCCACCCCA 813
                |||||||||:::|||||||||||||||||||||||||||||||||||||||||||
Db    241 IleHisArgLysGluAsnAlaGlyAlaAlaAlaGluLysProIleThrIleHisSerThrPro 260

Qy    814 GAGGGGACTTCTGAAGCATGCCCATGATTCTTGAAATCATGCGAGAAAGAGGCAGATGAG 873
                ||||||||||||||:::||||||:::||||||||||||||||||||||||||||||
Db    261 GluGlyCysSerAlaAlaCysLysIleIleMetGluIleMetGlnLysGlnAlaGlnAsp 280

Qy    874 ACCAAACTAGCCGGAGAGATTCCTCTGAAAATCTTGGCACACAATGGCTTGGTTGGAGA 933
                ||:|||||||||||||||||||||||||:::||||||||||||||||||||||||||
Db    281 ThrLysPheThrGluIleLeuProLeuLysIleLeuAlaHisAsnAsnPheValGlyArg 300

Qy    934 CTGATTGAAAGAGAAGGCAGAAATTTGAACAATGAAAATTGAACATGAAACAGGGACCAAGATA 993
                ||||||||||||||||||||||||||||||||||||||:::||||| :::|||
```

```
RESULT 4
O73932
ID   O73932                  PRELIMINARY;           PRT;         594 AA.
AC   O73932;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   VG1 RNA binding protein variant D.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Havin L., Git A., Elisha Z., Oberman F., Yaniv K.,
RA   Pressman Schwartz S., Standart N.M., Yisraeli J.K.;
RL   Genes Dev. 0:0-0(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98228351; PubMed=9560341;
RA   Deshler J.O., Highett M.I., Abramson T., Schnapp B.J.;
RT   "A highly conserved RNA-binding protein for cytoplasmic mRNA
RT   localization in vertebrates.";
RL   Curr. Biol. 8:489-496(1998).
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AF064634; AAC18598.1; -.
DR   EMBL; AF055923; AAC41285.1; -.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR005504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH TYPE 1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   594 AA;  65643 MW;  54CEA7BFF085SDD6 CRC64;
```

```
OC   Archosauria; Aves; Neognathae; Galliformes; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97220007; PubMed=9121465;
RA   Ross A.F., Oleynikov Y.S., Kislauskis E.H., Taneja K.L., Singer R.H.;
RT   "Characterization of a beta-actin mRNA zipcode-binding protein.";
RL   Mol. Cell. Biol. 17:2158-2165(1997).
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AF026527; AAB82295.1; -.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR005504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS00084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
SQ   SEQUENCE   576 AA;   63271 MW;   01AAF2D1D81C8811 CRC64;
```

```
Alignment Scores:
Pred. No.:                   1.29e-142      Length:            576
Score:                       2033.00        Matches:           402
Percent Similarity:          78.94%         Conservative:      74
Best Local Similarity:       66.67%         Mismatches:        95
Query Match:                 33.21%         Indels:            32
Db:                          13             Gaps:              6
```

```
US-09-270-437D-6 (1-3412) x O42254 (1-576)

Qy   73 ATGAACAAGCTTACATCGGAACTGAGCCCCGTCACCGGCCGAGACCTCCGGCAG 132
         |||||||||||||| :::    ::|||||||||||||| ::
Db    1 MetAsnLysLeuTyrIleGlyAsnLeuAsnGlueValThrProAlaAspLeuGluLys 20

Qy  133 CTCTTTGGGACGCAGGAGAGTGCCCGTGGCGGACAGGTCCTGCTGAAGTCCGGCTACGCC 192
         :::|||| || ||:||||||||   ::||||| ||||||||:: |||||
Db   21 ValPheAsnAspHisLysIleSerPheSerGlyGlnPheLeuValLysSerGlyTyrAla 40

Qy  193 TTCGTGGACTACCCCGACCAGAACTGGGCCATCGGCGCCATCGAGACCCTCGGGTAAA 252
         |||||||| ||:||||||||||| |||||||||||||||||||| :||||:||
Db   41 PheValAspCysProAspGluGlnTrpAlaMetLysAlaIleGluThrPheSerGlyLys 60

Qy  253 GTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAGCTAAGGAGC 312
         |||||||||||||||:: :||||| |||||||| ||||||:|||||||||||||
Db   61 ValGluLeuHisGlyLysIleMetGluLeuGluHisSerValProLysGlnArgSer 80

Qy  313 AGGAAAATTCAGATTGAAACTCCCTCCACCGTGGGAGGTGTTGGATGGACTT 372
         |||||||||||:|||||||||| :||||||||||||||:||:|||||
Db   81 ArgLysIleGlnIleArgAsnIleProProGlnLeuArgTrpGluValLeuAspGlyLeu 100

Qy  373 TTGGCTCAATATGGACCAGTGGACAGGTGAACAAGTCAACACAGACAGAACCGCC 432
         ||||:|||||||||||||||||||:| |||||| ::|||||||||||||||
Db  101 LeuAlaGlnTyrGlyThrValGluAsnCysGluGlnValAsnThrAspArgThrAla 120

Qy  433 GTTGTCAACGTCACATATGCACACAGAGAAGAAGCAAAAATAGCCATGGAGAAGGTAAGC 492
         |||||||||||:: :||||| :::||||  ||:|:|||||||:|||||
Db  121 ValValAsnValThrTyrThrAsnArgGluGlnThrArgGlnAlaIleMetLysLeuAsn 140

Qy  493 GGGCATCAGTTTGAGAACTACTCCTTCAAGATTCTACATCCCGATGAAGAGGTGAGC 552
         |||||||||:::    |||||:|::    ||||||::|
Db  141 GlyHisGlnLeuGluAsnHisValLeuLysValSerTyrIleProAspGluGln----- 158

Qy  553 TCCCCTTCAGCCAGCCACCTGGGGACCACTCTTCCGG------------GAG 603
         |||||||||||||||||:|||||| |||||:|||||
Db  159 SerValGlnLeuProGluAsnGlyArgGlyArgGlyPheGlyAlaProGlyAlaProArg 178

Qy  604 CAAGGCCAC------GCCCCTGGGGCGACACTTCCTCAGGCCAGCAGATTGATTTCCCGCTG 657
         |||||       ||| :::  :||||||:|  :::::|||||
Db  179 GlnGlySerProValThrAlaGlyAlaProValLysGlnProValLysGlnProIleProLeu 198
```

```
Dd  301 LeuIleGlyLysGluGlyArgAsnLeuLysLysIleGluGlnAspThrAspThrLysIle 320
Qy  994 ACAATCATCTTTGCAGGATTTGAGCATATACAACCCGAAAGAACCATCACTGTGAAG 1053
         ||||| :||||||||||||||| :::|||:||||||||||||||||||||||
Db  321 ThrIleSerProLeuGlnAspLeuThrLeuTyrAsnProGluArgThrIleThrValLys 340

Qy 1054 GGCCAGTTGAGGCCTGTGCCAGTGCAGTAGATAGAGATTATGAAGAAGCTGCGTGAGGCC 1113
         ||||:::|:::|||| |||| :::||||:::||||||:::||:: :||||
Db  341 GlySerIleGluProCysAlaAlaLysAlaGluGluGluMetLysLysIleArgGluSer 360

Qy 1114 TTTGAAAATGATATGTCTGGCTGGTGTTAACCAACAAGCCAATCTGATCCCAGGGTTGAACCTC 1173
         :::||||||::: ||:|||||| |||:||||  |:||| ::|||||
Db  361 TyrGluAsnAspIleAlaAlaMetAlaAlaHisLeuGlnAlaHisLeuIleProGlyLeuAsnLeu 380

Qy 1174 AGCCACCTTGGCATCTTTTCAACAGGACTGTCCGTGCTATCTCCACAGCAGCAGGGCCCCGC 1233
         ||: :::
Db  381 AsnAlaLeuGlyLeuPhe-----------------------ProSerSerSer 391

Qy 1234 GGAGCTCCCCCGCTGCCCCCTACCACCCCTTCACTACCACTCCGGATACTTCTCCAGC 1293
         || ||||||  :::||||||||| |||:: :||||||||||
Db  392 GlyMetProProSerValGlyValProSerProThrSer--------------SerThr 407

Qy 1294 CTGTACCCCCATCACCAGTTGGCCGTTCCCGCATCACTCTTATCCAGAGACGAGGAG 1353
         |||| ||||||
Db  408 SerTyrPro----------------------ProPheGlyGlnGln----------ProGluSerGlu 419

Qy 1354 ATTGTGAATCTCTTCATCCCAAGGTGTGGGCGCCATCATCGGGAAGAAGAGGGGCA 1413
         :::|||||| |||||||||||||||| || || ||||||||||||||||||
Db  420 ThrValHisLeuPheIleProAlaLeuAlaValAlaGlyAlaIleGlyAlaIleLeuIleGlyLysGlyGln 439

Qy 1414 CACATCAAACAGCTGGCGAGATTCGCCGAGCCCTCTATCAAGATGTGCCCTGCGGAGGC 1473
         |||||||||:::||||||||:|||:::|||:||:|||||||||||| :|:|||||||||
Db  440 HisIleLysGlnLeuSerArgPheAlaGluPheAlaGlyAlaSerIleLysMetCysProAlaGluAlaGlyGly 459

Qy 1474 CCAGACGTCAGCGAAAGGATGGTCATCATCACCGGGCACCGGAAGCCCAGTTCAAGGCC 1533
         ::|||||||||| |:::||||||||:::||||||||||||| ||||||||:||||||
Db  460 ProAspAlaLysLeuArgMetValIleIleThrGlyProProGlnAlaGlnPheLysAla 479

Qy 1534 CAGGGACGGATCTTCTTTGGGAAACTGAAAGAGGAAAACTTCTTTAACCCCAAAGAAGAGTG 1593
         ||||| |||:::|||||||||||||||:: |:::|||||||||||:::||||:||||||||
Db  480 GlnGlyArgIleTyrGlyLysLeuLysGluLeuAsnPheGlyLysGlnPheSerGlyGluVal 499

Qy 1594 AAGCTGGAAGCGGCATATCAGAGTGCCCTTCCACAGCTGGCCCGGTGATTGGCAAAGGT 1653
         |||||||||||||||||||:: :|||||:::||||||||| |||||||:::|:|||||||
Db  500 LysLeuGluThrHisIleLysIleLysLysValProSerTyrAlaAlaGlyValIleGlyLysGlyGly 519

Qy 1654 GGCAAGACCGTGAACGAACTGCAGAACTTAACCAGTGCAGAAGTGCATCGTGCCTCGTGAC 1713
         |||||:::||||||| |||| ||||||:::::|:::|||:||:::||||||:: |||||
Db  520 GlyLysThrValAsnGluLeuGlnAsnLeuThrSerAlaGluValValValProAsnValProArgAsp 539

Qy 1714 CAAAGCCCAGATGAAATGAAAATGAGGAAGAGTGATCGTCAGAAATATCGGCATGGTTGCTAGC 1773
         |:||:||||||||||||||||||||||:::||| |||||:||:: :|:||||
Db  540 GlnThrProAspGlyAsnAspGluAsnLysGlnLysValValValGlyValIleLysLysIleThrGlyLysHisPheTyrAlaSer 559

Qy 1774 CAGACTGCACAGCGCCAAGATCAGGGAGAATTGTACAACAGGTGAACGAGCAGCAGAAA 1833
         ||||:::||||||||||||||||:: :||||||:||||||||:::||||||||||||||
Db  560 GlnLeuAlaGlnArgLysIleGlnGlnLeuLeuLeuAlaGlnLysIleIleGlyValAlaArgArgGlnGlnGln 579

Qy 1834 TACCCTCAGGGAGTCGCCTCA 1854
         :: |||||||||||
Db  580 GlnLnLysThrValGlnSer 586
```

```
RESULT 5
O42254
ID   O42254        PRELIMINARY;    PRT;    576 AA.
AC   O42254;
DT   01-JAN-1998  (TrEMBLrel.  05,  Created)
DT   01-JAN-1998  (TrEMBLrel.  05,  Last  sequence  update)
DT   01-OCT-2003  (TrEMBLrel.  25,  Last  annotation  update)
DE   Zipcode-binding protein.
GN   ZBP1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
Qy  1738  GTGATCGTCAGAATTATCGGGCACTTCTTTGCTCAGCCAGACTGCACAGCCGCAAGATCAGG  1797
          |||||||||| |:|||||||| |||||||  |::||||||:|| |||||||||||:||||||
Db   534  ValIleValLysIleIleGlyHisPheTyrAlaSerGlnMetAlaGlnArgLysIleArg    553

Qy  1798  GAAATTGTACACAGGTGAAGCAGCAGCAGGAGAAATACCCTCAGGGAGTCGCCTCACAG    1857
          ::::|||:||||||||||||||||||||:|||
Db   554  AspIleLeuAlaGlnValLysGlnHisGlnLysGlyGlnSerGlyGlnLeuGlnLeuGlnAla  573

Qy  1858  CGGAGCAAG  1866
          |||   |||
Db   574  ArgArgLys  576
```

RESULT 6

```
ID    Q8CGX0          PRELIMINARY;    PRT;    577 AA.
AC    Q8CGX0;
DT    01-MAR-2003 (TrEMBLrel. 23, Created)
DT    01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT    01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE    B-actin zipcode binding protein 1.
OS    Rattus norvegicus (Rat).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX    NCBI_TaxID=10116;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=Sprague-Dawley;
RA    Eom T., Singer R.H., Bassell G.J.;
RT    "Molecular interactions between rZBP1 and b-actin zipcode required for
RT    transport of mRNA and stimulation of spine growth.";
RL    Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; AF541940; AAO16210.1; -.
DR    GO; GO:0003676; F:nucleic acid binding; IEA.
DR    InterPro; IPR004087; KH_dom.
DR    InterPro; IPR004088; KH_type_1.
DR    InterPro; IPR005504; RNA_rec_mot.
DR    Pfam; PF00013; KH; 4.
DR    Pfam; PF00076; rrm; 2.
DR    SMART; SM00322; KH; 4.
DR    SMART; SM00360; RRM; 2.
DR    PROSITE; PS50084; KH_TYPE_1; 4.
DR    PROSITE; PS50102; RRM; 2.
SQ    SEQUENCE  577 AA;  63436 MW;  0647676128FBD1EE CRC64;
```

Alignment Scores:

```
Pred. No.:                        6.59e-142      Length:            577
Score:                            2023.50        Matches:           395
Percent Similarity:               78.94%         Conservative:      81
Best Local Similarity:            65.51%         Mismatches:        96
Query Match:                      33.06%         Indels:            31
DB:                               577 AA;        Gaps:              4
```

US-09-270-437D-6 (1-3412) x Q8CGX0 (1-577)

```
Qy   73  ATGAACAAGCTTTACATCGGGACCTGAGCCGCCCGCGTCACCGGCGACGACCTCCGGCGCAG  132
         |||||||||||||||||||||||       :::
Db    1  MetAsnLysLeuTyrIleGlyThrAsnLeuAsnGluSerValThrProAlaAspLeuGluLys   20

Qy  133  CTCTTGGGACCAGGAGAGCTGCCCCTGCGGGACAGGTCCGCTGCTGAAGTCCGGCTACCGCC  192
         ::::||||| :::|||| |||||| ||||||||
Db   21  ValPheAlaGluHisGlnHisPheLeuValLysSerGlyTyrAla   40

Qy  193  TTCGTGGACTACCCCAGCAGCTGGGCCATCGGCATCGAGACCCTCCGGGTAAA  252
         |||||||| |||||  |||  |||||||| |||||||||:||||||||
Db   41  PheValAspCysProGlyAspAspValHisTrpAlaMetLysAlaIleGluThrPheSerGlyLys   60

Qy  253  GTGGAATTCATGGAGAAATCATGGAGAAATGATGATTACTCAGTCTCTAAAAAGCTAAGGAGC  312
         ||||||||||| |||  ||||||| ::||::|||||||
Db   61  ValGluLeuGlnLysGlyValLysLysThrValAsnGluLeuGln   80

Qy  313  AGGAAAAATCAGATTCGAAACATCCCTCCTCCACCTGCAGTGTGGGAGGTGTTGAGATGGACTT  372
         ||||||||||| :::||||||||||||
Db   81  ArgLysIleGlnIleArgAsnIleProProGlnLeuArgTrpGluValLeuAspArgAspLeu  100
```

Left column:

```
Qy   658  CGGATCCTGGTCCCCACCCAGTTGTTGTGTGCCATCATCGGAAAGGAGGGCTTGACCATA    717
          |||||||||||||  :::|||||||||||  ::::||||||||||||||
Db   199  ArgLeuLeuValProThrGlnTyrValGlyAlaIleIleGlyLysGluGlyLeuThrIleGlu  218

Qy   718  AAGAACATCACTAAGCAGCAGCCAGTCCCGGTAGTATCCATAGAAAGAGAGACTCTGGA    777
          :::|||||||||||||||||||::|||||||||  :::|||||||||||||||||
Db   219  ArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgLysSerGluAsnAlaGly  238

Qy   778  GCTGCAGAGAAGCCTGTCCACCATGCCAGCTCCCCAGAGGGACTTCTGAAGCATGCCGC    837
          ||||||||||||  :::||||    |||   ||| |||||||||||||:::
Db   239  AlaAlaGlyLysAlaIleAlaIleHisSerThrProGlyGlyCysSerAlaAlaCysLys    258

Qy   838  ATGAATTCTGAAATCATGCAGAAAGAGGCAGATGAGACAAACTAGCCGAAGAGATTCCT    897
          ||||||||||| |||||::|||||||||| ||||:::||||||| ||||||
Db   259  MetIleLeuGluIleMetGlnLysGluAlaAspThrLysThrAlaAspGluValPro    278

Qy   898  CTGAAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAGAAGAGGCAGAAAT    957
          |||||||||||||  :::|||||||||||
Db   279  LeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyArgGluGluGlyArgAsn   298

Qy   958  TTGAAGAAATTGAACATGAAACAGGGACCAAGATAACAATCCATCTTTGCGAGGATTTG   1017
          ||||||||||||||| |||  ::::||||
Db   299  LeuLysLysValGluGlnIleAspThrValThrLysIleThrIleSerSerLeuGlnAspLeu  318

Qy  1018  AGCATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTGAGGCCCTGCCAGT   1077
          :::||||||||||||||||||||  :::||||
Db   319  ThrLeuTyrAsnProGlnArgThrIleThrValLysGlyGlnSerGluAsnCysCysLys   338

Qy  1078  GCTGAGATAGAGATTATGAAGAGCTGCGTGAGGCCTTTGAAAATGATATGCTCGCTGTT  1137
          ||||  |||:::|||||||||||| |||  ||::::
Db   339  AlaGluGluGluIleMetLysLysValArgGluAlaTyrGluAsnAspValAlaAlaMet   358

Qy  1138  AACCAACAAGCCAATCTGATCCCCAGGGGTGAACCTCAGCGACTTGGCATCTTTTCAACA   1197
          :::||||||:::|||||||||||||
Db   359  SerLeuGlnSerHisLeuIleProGlyLeuAsnLeuAlaValGlyLeuPheProAla    378

Qy  1198  GGACTGTCGCGTGCTATCTCCACCAGCAGGGCGCCCCGCGAGCTCCCCCGCCTGCCCCTAC   1257
          |||||||||||  :::|||||||||||
Db   379  SerSerAsnAlaValProProPro-----ProSerValSerGlyAlaAlaProTyr    396

Qy  1258  CACCCCTTCACTACCACCTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGC   1317
          ||||
Db   397  ------------------------------------SerSerPheMetPro--------   401

Qy  1318  CCGGTTCCGCATCATCACTCTTATCACACGAGCAGGAGATTGTGAATCTCTTCATCCCAACC  1377
          |||||||  :::|||||
Db   402  -------------------------ProGluGlnGluThrValHisValPheIleProAla  413

Qy  1378  CAGGCTGTGGCGGCCATCATCGGAAGAGGGGGCCACACATCAAACAGTGTGGCCGAGATTC  1437
          :::|||||||||||||||||||  :::||||
Db   414  GlnAlaValGlyValGlyAlaIleIleGlyLysGlyGlyHisIleLysAsnLeuSerArgPhe  433

Qy  1438  GCCGGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCAGAGCGTCAGCGGAAGGATGTC   1497
          |||||||||||||||||||||  :::|||||
Db   434  AlaSerAlaSerIleLysIleAlaProValThrProSerLysSerValArgMetVal   453

Qy  1498  ATCATCACCGGGCCACCGGAAGCCCCAGTTCAAGGCCAGGGACGGATCTTTGGGAAAACTG   1557
          :::|||||||||||||||||||
Db   454  ValIleThrGlyProGluAlaGlnSerGlyLysGlnIleGlyArgIlePheGlyLysLeu   473

Qy  1558  AAAGAGGAAAAACTTCTTTAACCCCAAAGAAAGAAGTGAAGCGGCATATCAGAGTG   1617
          |||||||||||| :::||||||||
Db   474  LysGluGluAsnPhePheGlyProLysProLysGluLysGluLeuGluThrHisIleArgVal   493

Qy  1618  CCCTCTTCCACAGCTGCCGGGTGATTGGCAAGTGGCAAGGTGCAAGACCGTGAACGACTGCAG   1677
          :::|||||||||||||||||||
Db   494  ProAlaSerAlaAlaGlyArgValIleGlyLysGlyGlyLysThrValAsnGluLeuGln    513

Qy  1678  AACTTAACCAGTGCAGAAGTCATCGTGCCCTGTGACCAAACGCCAGATGAAAATGAGGAA   1737
          ||||||||| :::|||||||||||
Db   514  AsnLeuThrAlaAlaGluValIleValProArgAspAspGlnThrProAspGluAsnGluGln  533
```

```
Qy  373 TTGGCTCAATATGGGACAGTGGAGAATGTGGAGAACAAGTCAACACAGACAGAAAACCGCC 432
            ||||||||||||||||||||||||||||||||||:::||    ||||||||||||||
Db  101 LeuAlaGlnTyrGlyThrValGluAsnValGluAsnLysSerGlnThrGluGluAsnThrAla 120

Qy  433 GTTGTCAACGTCACATATGCAACACAGAGAGAAGCAAAAATAGCCATGAGAAGCTAAGC 492
            ||||||||||||||:::    ||| :::  ||||| ||||||||||  :::||:||
Db  121 ValValAsnValThrTyrSerAsnArgGluGlnThrArgGlnAsnAlaIleMetLysLeuAsn 140

Qy  493 GGGCATCAGTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGATGAAGAGGTGAGC 552
            ||||||::::||   ||||:::::::::  ::::::::::::
Db  141 GlyHisGlnLeuGluGlnAsnHisAlaLeuLysValSerTyrIlePro 160

Qy  553 TCCCCTTGCCCCCCTCAGCGAGCCCTGGGGACCCACTTCTCCCGGGAGCAAGGCCAC 612
            ||||::|   ||||||
Db  161 ------GlnGlyProGluAsnGlyArgArgGlyLeuGlyPheGlySerArgGlyGlnProArg 178

Qy  613 --------------GCCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTG 657
                          ||| |||     :::::   ||||  |||||||||
Db  179 GlnGlySerProValAlaAlaGlyAlaProAlaLysGlnGlnGlnAsnValAspSerIleProLeu 198

Qy  658 CGGATCCTGTCCCCACCCAGTTTGTTGTGCCATCATCGGAAGGAGGGCCTTGACCATA 717
            ||||||||||||||||:::||||||||| ||||||||||||||||
Db  199 ArgLeuLeuValProThrHisGlnTyrValGlyAlaIleIleGluGlyLysGlyAlaThrIle 218

Qy  718 AAGAACACTAAGCACGAGAGCCCAGTCCCGGGTAGAATCCCATAGAAAAGAAGAACTCTGGA 777
            |||||||||||||||||||||:::|||||||  ::::::::::::||||||||||||:||
Db  219 ArgAsnIleThrLysSerGlnThrGlnSerLysIleAspThrHisArgLysValGluAsnAlaGly 238

Qy  778 GCTGCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGC 837
            |||||||||||||||:::|||||||||||||||||||||   :::
Db  239 AlaAlaLysAlaAlaLysSerAlaIleSerAlaValHisSerThrProGluGluGlyGlyCysSerSerAlaCysLys 258

Qy  838 ATGATTCTTGAAATCATGCAAGAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCT 897
            ||||||||||||||||||||||||||||||||||||||||
Db  259 MetIleLeuGluIleMetHisGluGlyAspGluThrLysLeuAlaGluGluIlePro 278

Qy  898 CTGAAAATCTTGGCACACAATGGCTTGGTTGGAAGCTGATTCGGAAAAGAAGCCAGAGAAT 957
            ||||||||||:::|||  ||||||||||  |||||||||||||
Db  279 LeuLysIleLeuAlaHisAsnAsnPheValGlyLeuLeuGlyLysLeuIleArgLysGluGlyGlyArgAsn 298

Qy  958 TTGAAGAAATTGAACATGAACATGGGACGACCAAGATAACAATCTCATCTTTGCCAGGATTTG 1017
            |||||||:::||       ||||||||||||||||||||
Db  299 LeuLysLysValGlyGluGlnAspThrGluThrLysHisIleThrIleSerSerLeuGlnAspLeu 318

Qy 1018 AGCATATACACCCGGAAGAACCCATCACTGTGAAGGGCCACAGTTGAGGCCTGTGCCAGT 1077
            :::::|||||||||||  |||||||  :::||
Db  319 ThrLeuTyrAsnProGluArgThrIleThrValLysGlyAlaIleGluLysLeuAsnCysCysSerArg 338

Qy 1078 GCTCGAGATAGAGATTATGAAGAAGAGCTGCTGAGGCCTTTGAAAATGATATGTCTGGCTGTT 1137
            |||||||||||||||||||||:::||||||||||||||||||
Db  339 AlaGluGlnGluIleMetLysLysMetLysValAsnArgGluAlaTyrGluAlaValAlaAlaAlaMet 358

Qy 1138 AACCAACAAGCCAATCTGATCCCCAGGGTTGAACCTCAGCGCACTTGCATCTTTTCAACA 1197
            :::    ||:|||||||||||||||||||||||
Db  359 SerLeuGlnSerHisSerLeuIleProGlyLeuAsnLeuAlaAlaValAlaGlyLeuPheProAla 378

Qy 1198 GGACTTCCGTGTCATCCCACCAGCAGGAGTCTCCCCGCGCCCCCTAC 1257
            ||||||    :::||||
Db  379 SerSerSerAlaValProProPro------ProSerSerValThrGlyValAlaAlaProTyr 396

Qy 1258 CACCCCTTCACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTTGGC 1317
            :::   ||:||||||
Db  397 GlySerPheMet------ 400

Qy 1318 CCGTTCCCGCCATCATCACTCTTATCCAGACGACGAGATTGTGAATCCTTCATCCCAACC 1377
            |||||  :::||||
Db  401 ------GlnAlaProGluGlnGluMetValGlnMetValGlnHisIleLysLysLeuSerGlnArgPhe 414

Qy 1378 CAGGCTGTGGGGGGCCCATCATCGGGAAGAGGGGCCACATCAAACAGTCTGGCGAGATTC 1437
            |||||||||||||||||||||||||||||||||||:::||
Db  415 GlnAlaValAlaGlyValAlaIleIleIleGlyLysLysGlyGlyHisIleLysLysIleArgPhe 434
```

```
Qy 1438 GCCGGAGCCTCTATCAGATTGCCCCTGCGGAAGGCCCAGAGTCAGCGAAGGATGGTC 1497
            |||    ||||||||||||:::|||  |||||||  ||| |||||||
Db  435 AlaSerAlaSerIleLysIleAlaProProGluThrProAspSerLysValArgMetVal 454

Qy 1498 ATCATCACCGGCGCCACCGGAGCCCAGTTCAAGGCCCAGGACGATCTTTGGGAAACTG 1557
            :::|||||||||||||||||||||||||:::|||||| ||:::||||||
Db  455 ValIleThrGlyProProGluAlaGlnPheLysAlaGlnAsnGlyArgIleTyrGlyLysLeu 474

Qy 1558 AAAGAGGAAAACTTCTTAACCCAAGAAGAAGAGTGGAAGCGCATATCAGAGTG 1617
            |||||||||||||||||||||||||||||||||||||||
Db  475 LysGluGluAsnPhePheGlyProLysGluGluValLysLeuGluThrHisIleArgVal 494

Qy 1618 CCCTCTTCCACAGCTGCCGCGGTGATTGGCAAGAGTGGCAAGACCGTGAACGAACTGCAG 1677
            ||||||||||||||||||||||||||||||||||||||||||
Db  495 ProAlaSerAlaAlaGlyTyrArgValIleleGlyLysSerGlyGlyLysThrValAsnGluGln 514

Qy 1678 AACTTAACCAGTGCAGAAGTCATCGTGCCCCTGCTGACCAAACGCCAGATGAAAATGAGGAA 1737
            ||||||||||||||||||||||||||||||||||||||||||
Db  515 AsnLeuThrAlaAlaGluValValProLeuLeuProArgAspGlnThrProAspGluAsnAspGln 534

Qy 1738 GTGATCGTCAGAATTATCGGGACACTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGG 1797
            ||||||||||||||||||||||||||||||||||||||||||
Db  535 ValIleValLysIleIleGlyThrLeuPheAlaSerGlnThrAlaGlnArgLysIleArg 554

Qy 1798 GAAATTGTACAACAGTGAAGCAGCAGAAATACCCTCAGGGAGTGCGCTCCACAG 1857
            :::||||  :::||||||||||||||||||:::
Db  555 AspIleLeuGlnGlnValLysGlnGlnLysTyrProGlnGlyValArgLeuGluAlaGlnAla 574

Qy 1858 CGCAGCAAG 1866
            |||||
Db  575 ArgArgLys 577
```

RESULT 7
O88477
ID   O88477   PRELIMINARY;   PRT;   577 AA.
AC   O88477;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Coding region determinant binding protein.
GN   IGF2BP1 OR CRDBP.
OS   Mus musculus (Mouse)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92217743; PubMed=1559612;
RA   Bernstein P.L., Herrick D.J., Prokipcak R.D., Ross J.;
RT   "Control of c-myc mRNA half-life in vitro by a protein capable of
RT   binding to a coding region stability determinant.";
RL   Genes Dev. 6:642-654(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94158886; PubMed=8114742;
RA   Herrick D.J., Ross J.;
RT   "The half-life of c-myc mRNA in growing and serum-stimulated cells:
RT   influence of the coding and 3' untranslated regions and role of
RT   ribosome translocation.";
RL   Mol. Cell. Biol. 14:2119-2128(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94179348; PubMed=8132663;
RA   Prokipcak R.D., Herrick D.J., Ross J.;
RT   "Purification and properties of a protein that binds to the C-terminal
RT   coding region of human c-myc mRNA.";
RL   J. Biol. Chem. 269:9261-9269(1994).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97322234; PubMed=9178888;
RA   Leeds P., Kren B.T., Boylan J.M., Betz N.A., Steer C.J.,
RA   Gruppuso P.A., Ross J.;
RT   "Developmental regulation of CRD-BP, an RNA-binding protein that

RT   stabilizes c-myc mRNA in vitro.";
RL   Oncogene 14:1279-1286(1997).
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Ross J., Prokipcak R.D., Leeds P., Doyle G.A.R., Betz N.A.,
RA   Fleisig A.J.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryo;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K.H., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AF061569; AAC72743.1; -.
DR   EMBL; AK013940; BAB29071.1; -.
DR   HSSP; P11940; 1CVJ.
DR   MGD; MGI:1890057; Igf2bp1.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   577 AA; 63450 MW; EFFB1AF2FF9F0344 CRC64;

Alignment Scores:
Pred. No.:   9.29e-142
Score: 2021.50   Length: 577
Percent Similarity: 78.94%   Matches: 394
Best Local Similarity: 65.34%   Conservative: 82
Query Match: 33.03%   Mismatches: 96
DB: 11   Indels: 31
                                                         Gaps: 4

US-09-270-437D-6 (1-3412) x O88477 (1-577)

Qy   73 ATGAACAAGCTTTACATCGGGAACCTGGAGCCCGTCACCGCGCGACACCTCCGGCAG 132
         |||||||||||||||||||||||||||||||| :::||||||||||||||| :::
Db   1 MetAsnLysLeuTyrIleGlyAsnLeuGluSerValThrProAlaAspLeuGluLys 20

Qy   133 CTCTTTGGGGACAGGAAGTGCCCTCCGGGGGAGCAGGTTCCTGCTGAAGGTGGCTACGCC 192
         ::||||| || ||| :::||||| |||||| ||| ||||:|||||||||||
Db   21 ValPheAlaGluHisLysIleSerTyrSerGlyInPheLeuValLysSerGlyTyrAla 40

Qy   193 TTCGTGGACTACCCCACCAGAACTGGGCACTCAGCCCGCCATCGAGACCCTCCGGGTAAA 252
         ||||||||||||| ||||||||| ||| ||||||||||||||||||||||||||||||
Db   41 PheValAspCysProAspGluHisTrpAlaMetLysAlaIleGluThrPheSerGlyLys 60

Qy   253 GTGGAATTGCATGGGAGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGGAGC 312
         ||||||||| ||||| ||||||||||||||||||||:||||| :: |||||||||||||||
Db   61 ValGluLeuGlnGlyLysArgLeuGluIleAspTyrSerValSerLysLysLeuHisSerGlnArgSer 80

```
Pred. No.:                      2.6e-141        Length:          577
Score:                          2015.50         Matches:         394
Percent Similarity:             78.61%          Conservative:     80
Best Local Similarity:          65.34%          Mismatches:       98
Query Match:                    32.93%          Indels:           31
Db:                                 4           Gaps:              4

US-09-270-437D-6 (1-3412) x Q9NZI8 (1-577)

Qy    73 ATGAACAAGCTTTACATCGGGAACCTGAGCCCGCGCCCGGTCACCGACGCCTCCGGCAG 132
         :::|||||||||||||||||||||||||||||||||||||||||||||| ||||    ::
Db     1 MetAsnLysLeuTyrIleGlyAsnLeuSerValThrProAlaAspLeuLysGluLys 20

Qy   133 CTCTTTGGGACGAGGAGGCTGCCCCTGGGCGGACAGGTCCTGCTGAAGTCCGGCTACGCC 192
         ::|||     |||||  :: |||||||||||||||  :::||||
Db    21 ValPheAlaGluHisLysIleLysIlePheLeuValLysSerGlyTyrAla 40

Qy   193 TTCGTGGACTACCCCGACCAGACTGGGCCATCGGGCCCCATCGAGACCCTCTCGGGTAAA 252
         ||||||  |||||   |||||||||||||||||||||
Db    41 PheValAspCysProAspGluHisThrTrpAlaMetLysAlaIleGluThrPheSerGlyLys 60

Qy   253 GTGGAATTGCATGGGAAAATCATGAAGAGTTGATTACTCAGTCTCTAAAAGCTAAGGAGC 312
         ||||| |||||||||||||||||||||||  :::::||||||||||||
Db    61 ValGluLeuGlnGlyLysArgLeuLeuGluHisSerValProLysGlnHisSerGlnArgSer 80

Qy   313 AGGAAAATTCAGATTGCAAACATCCCTCCTCACCTGCAGTGGGAGGTGTTGATGACTT 372
         ||||||||||||||||||||||||||||  ::||||||||
Db    81 ArgLysIleGlnIleGlnAspIleProProGlnLeuArgTrpGluValLeuLeuAspSerLeu 100

Qy   373 TTGGCTCAATATGGACAGTGGAGATGTGGAACAAGTCAACACAGACACAGAAACCGCC 432
         ||||||  ||||||||||||||||||||  :::|||
Db   101 LeuAlaGlnTyrGlyThrValGluAsnCysGluGlnValLysAsnThrGluSerGluThrAla 120

Qy   433 GTTGTCAACGTCACATATGCACAAGAGGAAGCAAAAATAGCCATGGAAGCTAAGC 492
         |||||||||||||  :: :::  :::||||||||||||
Db   121 ValValAsnValThrTyrSerAsnArgGluGlnThrArgGlnAlaIleMetLysLeuAsn 140

Qy   493 GGGCATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCCGATGAAGAGGTGAGC 552
         ||||||||  :::|||||||  ||||||||
Db   141 GlyHisGlnLeuGlnAsnHisAlaLeuLysValSerTyrIleProAspGlnIleAla 160

Qy   553 TCCCCTTGCCCCCCAGCGAGCCCCAGTCGGGGGACCACTCTTCCCGGGGAGCAAGGCCAC 612
         |||||||||||||||
Db   161 ------GlnGlyProGluAsnGlyArgArgArgGlyGlyGlyPheGlySerArgGlyGlnProArg 178

Qy   613 -------------GCCCCTGGGGGACCACTTCTCAGGCACCAGACAGATTGATTTCCCGCTG 657
                      |||  :::::::|||||||
Db   179 GlnGlySerProValAlaAlaGlyAlaProAlaAlaLysGlnGlnGlnValAspIleProLeu 198

Qy   658 CGGATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCCGGAAGGAGGGGCTTGACCATA 717
         ||||||  :::::||  :::||||||||||||||||||
Db   199 ArgLeuLeuValProThrGlnTyrValGlyValAlaIleIleGluGlyLysAlaThrIle 218

Qy   718 AAGAACATCACTAAGCAGACCCAGTCCCGGTAGATATCCAGTAGAAAAGAGAACTCTGGA 777
         ::|||||||||||||||||||  :::::||||||
Db   219 ArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgLysGluAsnAlaGly 238

Qy   778 GCTGCAGAGAGCCTGTCACCATCCATGCCACCCCCAGAGGGGGACTTCTGAAGCATGCCGC 837
         ||||||||||  :::::::|||||||||||||||||||
Db   239 AlaAlaGluGluLysAlaIleSerValHisSerThrProGluGlyCysSerSerAlaCysLys 258

Qy   838 ATGATTCTTGAAATCATGCAAGAAGAGGCAGATGAGACCAAACTAGCCGAAGAATTCCT 897
         ::|||||||||||||  ::||||||||||||||
Db   259 MetIleLeuGluIleMetHisLysGluAlaLysLysAspThrLysThrAlaAspGluValPro 278

Qy   898 CTGAAAATCTTGGCACCAATGGCCTTGGTTGGAGACTGATTGGAAAAGGAGGAGCAGAAAT 957
         :::||||||||||||||||||||||
Db   279 LeuLysThrLeuAlaHisAsnAsnPheValGlyArgLeuIleGluGlyLysGlyArgGlnAsn 298

Qy   958 TTGAAGAAATTGAACATGAACATGAAAGAGGACCAAGATAACAATTCATCTTTGCAGGATTTG 1017
         |||||:::::||  ::|||
Db   299 LeuLysLysValGluGlnLysAspThrArgGluThrLysLysIleThrLysSerLeuAsnAspLeu 318

Qy  1378 CAGGCTGTGTGGGGCGCCATCATCGGGAAGAGAGGGGGCACACATCAAACAGCTGGCGAGATTC 1437
         |||||||||||||||||||||||||||||||||||||||||||::||||
Db   415 GlnAlaValGlyAlaIleIleGlyLysLysGlyGlnHisIleLysGlnLeuLysSerGlnArgPhe 434

Qy  1438 GCCGAGCCCTCTATCAAGATTGCCCTGCCGGGAAGCGTCAGCAGAAAGGATGGTC 1497
         |||||||||  ::|||||||||||||||||||||||||||||||||
Db   435 AlaSerAlaSerLeuLysAlaSerIleAlaProProGluThrProAspSerLysValArgMetVal 454

Qy  1498 ATCATCACCGGGGCCACCGGACCCAGTTCAAGGCACCGGACGGATCTTTGGGAAACTG 1557
         :::||||||||||||||||||||||||||||||||||||:::||||
Db   455 ValIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysLeu 474

Qy  1558 AAAGAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGAAGCCGCATATCAGAGTG 1617
         ||||||||  :::|||||||||||||  ||||||||||||||||||
Db   475 LysGluGluAsnPhePheGlyProLysGluGluValLysLeuLysProHisIleIleArgVal 494

Qy  1618 CCCTCTTCCACAGGTGCCTGGCCGGGTGATTGGCAAAGGTGGCAAAGACCGTGAACGAACTGCAG 1677
         ||  :::||||||||||||||||||||||||||||||||
Db   495 ProAlaSerAlaAlaGlyArgValIleGlyLysGlyGlyLysThrValAsnGluLeuGln 514

Qy  1678 AACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAA 1737
         |||||||||  :::||||||||||||  :::::|||||
Db   515 AsnLeuThrAlaAlaGluValIleValProArgAspGlnThrProAspGluAsnAspGln 534

Qy  1738 GTGATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGG 1797
         ||||||||||||||  :::|||||||||||||||||||
Db   535 ValIleValIleIleGlyHisPheTyrAlaSerGlnThrAlaGlnArgLysIleArg 554

Qy  1798 GAAATTGTACAACAGGTGAAGCAGCAGAGCAGAAATACCCTCAGGGAGTCGCCTCACAG 1857
         ::::::||  :::||||
Db   555 AspIleLeuAlaGlnValLysGlnGlnSerArgAsnLeuAlaGlnAla 574

Qy  1858 CGCAGCAAG 1866
         |||  |||
Db   575 ArgArgLys 577

RESULT 8
Q9NZI8
ID   Q9NZI8          PRELIMINARY;      PRT;     577 AA.
AC   Q9NZI8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   mRNA-binding protein CRDBP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ioannidis P., Trangas T., Dimitriadis E., Samiotaki M.,
RA   Panoutsakopoulos G., Kyriazoglou I., Voutoulias S., Tsiapalis C.M.,
RA   Kittas C., Agnantis N., Pandis N.;
RT   "Ectopic expression of a KH-domain containing protein, highly
RT   homologous to both human IMP-1 and mouse CRD-BP, in benign and
RT   malignant mesenchymal tumors.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AF198254; AAF37203.1; -.
DR   HSSP; P11940; 1CVJ.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   577 AA;  63456 MW;  0749A060F252D81D CRC64;

Alignment Scores:
```

[1]
RN
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RX   MEDLINE=22356683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573 (2002).
DR   EMBL; AK044850; BAC32119.1; -.
DR   MGD; MGI:1890357; Igf2bp1.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   577 AA;   63378 MW;   D75897450841532E CRC64;

Alignment Scores:
Pred. No.:   1
Score: 2014.50
Percent Similarity: 78.77%
Best Local Similarity: 65.17%
Query Match: 32.91%
DB: 11

US-09-270-437D-6 (1-3412) x Q8BRH1 (1-577)

```
Qy    73 ATGAACAAGCTTTACATGGGAACCTGAGCCCGACCGCCGTCACCGGCGGCCCCGACGACTTCCGGCAG 132
          ||||| :::::::||||||||||| :::::|| ||   |||||     ||||| ::::: ||:|||||
Db     1 MetAsnLysLeuTyrIleGlyAsnLeuAsnSerValThrProAlaAspLeuGluLysGlu 20

Qy   133 CTCTTGGGACAGGAAGCTGCCCTGCCCGGGACAGGTCCGCGGGCAGGTCCGGCTCCGGCTACGCC 192
          :::||| || ::|        ::: ||||||||||||::|||||||| |||
Db    21 ValPheAlaGluHisLysIleSerTyrSerGlyTyrGlyPheLeuValLysSerGlyTyrAla 40

Qy   193 TTCGTGGACTACCCCGACCAGAACTGGGCCATCCGGCACTGGAGACCCTCCGGGTAAA 252
          ||||||||| |||||:::::::||| |::||:::|:||||||||::|||||||||||
Db    41 PheValAspCysProAspHisAlaThrMetLysAlaAlaIleGluHisPheSerGlyLys 60

Qy   253 GTGGAATTGCATGGGAAAATCATGGAAGTGATTACTCAGTCTCTAAAAAGTAAGGAGC 312
          |||||| |||||||||||||||| ::||:|| |:|:|:|:|||| :::::||:|||||
Db    61 ValGluLeuGlnGlyLysArgLeuGluIleGluHisSerValProLysLysGlnArgSer 80

Qy   313 AGGAAAATTCAGATTCGAAACATCCTCCTCACCTGCCAGTGGAGGTGTTGGATGGACTT 372
          |||||||||||||| ||||||:::|:::||||||||:||:||:||||| |:::|||||
Db    81 ArgLysIleGlnIleArgAsnIleProProGlnLeuArgTrpLeuValLeuAspSerLeu 100

Qy   373 TTGGCTCAATATGGGACAGTGGAGAATGTGGAGAACTCAACACAGACAAAACCCGCC 432
          ||:::::||||:||||::|| ||||||||||| ||:::::|| |:|:::::|||||
Db   101 LeuAlaGlnTyrGlyThrValGluAsnCysGluValAsnThrGluSerGluThrAla 120

Qy   433 GTTGTCAACGTCACATATGCAACAAGAGAAGAGAGCAAAAATAGCCATGGAGAAGTAAGC 492
          ||:|||:::|| ||  :: ||  :::||||||  |||:|:: || ::|||:::::|||
Db   121 ValValAsnValThrTyrSerAsnArgGlnThrArgGlnAlaIleMetLysLeuAsn 140

Qy   493 GGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGATGAAGAGGTGAGC 552
          ||||||||||||||||:::::::||:::::|||  ||| ||:::||:::||||:|||
Db   141 GlyHisGlnLeuGluAsnHisAlaLeuLysValSerTyrIleProAspGluGlnIleThr 160

Qy   553 TCCCCTTGCCCCCTGCCCAGGCGAGACCACTCTTCCCGGGACCAAGGCCAC 612
          |::::::||   ||||:::::||||:::::||||||||:::||:||||:||
Db   161 ------GlnGlyProGluAsnGlyAspArgArgGlyPheGlySerGlyGlnProArg 178

Qy   613 --------------------------GCCCCTGGGGGCACTTCTCAGGCGCCAGACAGATGATTTCCGCTG 657
                                      ||:||    ::: :::   ||||||
Db   179 GlnGlySerProValAlaAlaGlyAlaProAlaLysGlnGlnProValAspProLeu 198
```

```
Db    535 VallleValLysIleIleGlyHisPheTyrAlaSerGlnMetAlaGlnArgLysIleArg  554
              |||||||||||||||||::|||||||||||||||||||

Qy   1798 GAAATTGTACAACAGGTGAAGCAGTGAAGCAGCAGGAGCAGAAATACCTCAGGGAGTCGCC CACAG 1857
              :::||||:::  ||||| |||||||    ::||
Db    555 AspIleLeuAlaGlnValLysGlnHisGlnLysGlyGlnSerAsnLeuAlaGlnInAla  574

Qy   1858 CCGCAGCAAG 1866
              ||  |||
Db    575 ArgArgLys  577
```

RESULT 10
Q80US9

```
ID   Q80US9              PRELIMINARY;      PRT;     577 AA.
AC   Q80US9;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Igf2bp1 protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Sv x 129/Sv-CP; TISSUE=Embryonic stem cells;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Sv x 129/Sv-CP; TISSUE=Embryonic stem cells;
RA   Strausberg R.;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC051679; AAH51679.1; -.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE 577 AA; 63378 MW; D439F544257DA3CE CRC64;
```

Alignment Scores:
```
Pred. No.:                            Length:             577
Score:                   3.09e-141    Matches:            393
                         2014.50      Conservative:        82
Percent Similarity:      78.77%       Mismatches:          97
Best Local Similarity:   65.17%       Indels:              31
Query Match:             32.91%       Gaps:                 4
DB:                      11
```

```
Qy    658 CGGATCCTGGTCCCCACCCAGTTCTTGGTGCCATCATCGGAAAGAGAGGGCTTGACCATA 717
              |||::::|||||:::|||||||   |||||
Db    199 ArgLeuLeuValProThrGlnTyrValGlyAlaIleIleGlyLysGluGlyAlaThrIle 218

Qy    718 AAGAACATCACTAAGCAGACAGACCCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGGA 777
              |||||||||||| :::|||  |||||| ||||||
Db    219 ArgAsnIleThrLysGlnThrGlnSerValIleAspValHisArgLysGluAsnAlaGly 238

Qy    778 GCTGAGAGAAGCCTGTCACCATCATGCCACCCCAGAGGGACTTCTGAAGCATGCCGC 837
              ||||| :::::::||| ||| |||  ||::::
Db    239 AlaAlaGluLysAlaIleSerValHisSerThrProGluGlyCysSerSerAlaCysLys 258

Qy    838 ATGATTCTTGAAATCATGCAGAAAGAGCCAGATGAGACCAAACTAGCCGAAGAGATTCCT 897
              |||||||||||||||:::||
Db    259 MetIleLeuGluIleMetHisLysLysLysAlaAlaLysAspThrLysThrAlaGluValPro 278

Qy    898 CTGAAAAATCTGGCACACAATGGCTTGGTTGGAGAACTGATTGGAAAAGAAGGCAGAAAT 957
              |||||||||||||| |||||||| |||||||
Db    279 LeuLysIleLeuAlaHisAsnGlyLeuValGlyGluLeuIleGlyLysGluGlyArgAsn 298

Qy    958 TTGAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCCATCTTTGCAGGATTTG 1017
              ||||||:::|||   ::::|||   ||||
Db    299 LeuLysLysValGluGluGlnThrGluThrThrIleThrIleSerSerLeuGlnAspLeu 318

Qy   1018 AGCATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGCCTGTGCCAGT 1077
              ::::|||||| |||||||| ||||||||||
Db    319 ThrLeuTyrAsnProGluArgThrIleThrValLysGlyAlaIleGluAsnCysCysArg 338

Qy   1078 GCTGAGATAGAGATTATGAAGAGCTGCCTGAGCGCCTTTGAAAATGATATGTCTGGCTGTT 1137
              ||||||||| ||||||||||
Db    339 AlaGluGluGlnIleGluMetLysSerLysValArgGluIleLeuThrAsnAspValAlaAlaMet 358

Qy   1138 AACCAACAAGCCAATCGATCCCCAGGGTGAACCTCAGCGCACTTGGCCATCTTTTCAACA 1197
              :::  ||||||:::||||  ||::::||
Db    359 SerLeuGlnSerHisLeuLeuProGlyLeuAsnLeuAlaValGlyLeuPheProAla 378

Qy   1198 GGACTGTCCGTGCTATCTCCACCAGCAGGGCCGAGCTCCCCCGTCGCCCCCTAC 1257
              :::  |||||     ||||||||||||
Db    379 SerSerSerAlaValProProPro------ProSerSerValThrGlyAlaAlaProTyr 396

Qy   1258 CACCCCTTCACTACCCCACTCGGGATACTTCTCCAGCTGTACCCCCATCACCAGTTTGGC 1317
              |||     |||||:::|||||
Db    397 SerSerPheMet-------------------------------------------------- 400

Qy   1318 CCGTTCCCGCCATCATCACTCTATCCAGCAGGAGATTGTGAATCTCTTCATCCCAACC 1377
              ||||||||   :::||   ||||:::|||
Db    401 ---------------------GlnAlaProGluGlnInGlyMetValGlnValPheIleProAla 414

Qy   1378 CAGGCTGTGGGCGCCATCATCGGGAAGAGAGGGGGCCACACATCAAACAGCTGCGCGAGATTC 1437
              ||||||||| ||||||||||||
Db    415 GlnAlaValGlyIleAlaIleIleGlyLysGlyGlnInHisIleLysSerGlnInLeuSerArgPhe 434

Qy   1438 GCCGGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGAGCGTCAGCGAAAGGATGGTC 1497
              |||||||| |||||||:::|||  ||::::||
Db    435 AlaSerAlaSerLeuIleLysIleAlaProProGluThrProAspSerLysValArgMetVal 454

Qy   1498 ATCATCACCGGCCACGGGAAGCCCCAGTTCAAGGCGACGGATCTTTGGGAAACTG 1557
              :::||||:::|||  |||||:::||
Db    455 ValIleThrGlyProGlyGlnAlaGlnPheLysAlaGlnGlnArgIleTyrGlyLysLysLeu 474

Qy   1558 AAAGAGGAAAACTTCTTTAACCCCAAAGAAGAGAAGTGAAGCTGAAGGCCATATCAGAGTG 1617
              ||||||||||||:::|||
Db    475 LysGluGluGluAsnPhePheGlyProLysLysGluGluLysLeuGluThrHisIleArgVal 494

Qy   1618 CCCTCTTCCACAGCTGGCGCGGGTGATTGGCAAAGTGCAAGACCGTGAACGAACTGCAG 1677
              ||||||:::||||||||||
Db    495 ProAlaSerAlaAlaGlyValIleGlyValGlyLysGlyIleLysGluIleLeuGln 514

Qy   1678 AACTTAACCAGTGCCAGAAGTCACGTGCCTCGTGACCAAACGCCAGATGAAATGAGGAA 1737
              ||||||  |||||||:::|||||||
Db    515 AsnLeuThrAlaAlaGluValValValProAlaArgAspGlnThrProGluAsnAspGln 534

Qy   1738 GTGATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGATGCACAGCGCAAGATCAGG 1797
```

US-09-270-437D-6 (1-3412) x Q80US9 (1-577)

```
RESULT 11
Q9PW80
ID   Q9PW80        PRELIMINARY;      PRT;     582 AA.
AC   Q9PW80;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Vg1 RNA binding protein (Decapentaplegic and Vg-related 1, RNA binding
DE   protein).
GN   DVR1RBP.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Zhang Q., Yaniv K., Oberman F., Wolke U., Git A., Fromer M.,
RA   Taylor W., Meyer D., Standart N., Raz E., Yisraeli J.K.;
RT   "Vg1 RBP intracellular distribution and evolutionarily conserved
```

```
RT    expression suggest multiple roles during development.";
RL    Mech. Dev. 0:0-0(1999).
RN    [2]
RP    SEQUENCE FROM N.A.
RC    STRAIN=AB; TISSUE=Body;
RX    MEDLINE=22388257; PubMed=12477932;
RA    Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA    Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA    Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA    Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA    Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA    Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA    Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA    Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA    Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA    Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA    Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA    Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA    Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA    Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA    Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA    Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA    Jones S.J., Marra M.A.;
RT    "Generation and initial analysis of more than 15,000 full-length human
RT    and mouse cDNA sequences";
RL    Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN    [3]
RP    SEQUENCE FROM N.A.
RC    STRAIN=AB; TISSUE=Body;
RA    Strausberg R.;
RL    Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
CC    -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR    EMBL; AF161270; AAD45610.1; -.
DR    EMBL; BC045873; AAH45873.1; -.
DR    ZFIN; ZDB-GENE-000308-1; dvr1rbp.
DR    GO; GO:0003676; F:nucleic acid binding; IEA.
DR    InterPro; IPR004087; KH dom.
DR    InterPro; IPR004088; KH_type_1.
DR    InterPro; IPR000504; RNA_rec_mot.
DR    Pfam; PF00013; KH; 4.
DR    Pfam; PF00076; rrm; 2.
DR    SMART; SM00322; KH; 4.
DR    SMART; SM00360; RRM; 2.
DR    PROSITE; PS50084; KH_TYPE_1; 4.
DR    PROSITE; PS50102; RRM; 2.
SQ    SEQUENCE   582 AA;  63351 MW;  9DAE63200681B306 CRC64;
```

Alignment Scores:

```
Pred. No.:                          Length:           582
Score:              2.29e-138       Matches:          393
                    1976.00         Conservative:      75
Percent Similarity: 79.19%          Mismatches:       101
Best Local Similarity: 66.50%       Indels:            22
Query Match:        32.28%          Gaps:               4
DB:                 13
```

```
US-09-270-437D-6 (1-3412) x Q9PW80 (1-582)

Qy    73 ATGAACAAGCTTTACATGGGAACCTGAGCCCGTCACCCGTCACCCGACGACCTTCCGCGCAG 132
          ||||||||||||||:::||||  ||||    :::||| ||||    |||||
Db     1 MetAsnLysLeuTyrIleGlyAsnValSerGluGlnAlaSerAlaLeuAspLeuGluSer  20

Qy   133 CTCTTTGGGGACAGGAAGTGCCCCTGCGGGGACAGGTCCTGTGAAGTCCGGCTAGCGC    192
          ::|||||  ||| |||:::|||||    ||| ||||||||  ||:::||
Db    21 IlePheGluGlnTrpLysIleProPheSerAlaProPheLeuValLysSerGlyTyrAla  40

Qy   193 TTCGTGGACTACCCCGACCAGAACTGGGCCATCCGCGCCATCCGAGACCCCTCTCGGGTAAA 252
          ||||||||||:::||  |||||||  ||| ||||  |||||||||:||
Db    41 PheValAspCysProArgGlnAsnTrpAlaIleArgAlaIleArgAspThrLeuSerGlyLys  60

Qy   253 GTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGGAGC    312
          |||||:::||||| |||||||||:::||||||
Db    61 ValGluGluHisGlyLysValLeuGluValHisGluHisGlyLysGlyGlnAlaHisGluAla  80

Qy  1387 GGCGCCATCGACCATGCGGAAGAAGAGGGGCCACACATCAAACAGTCTGGCGGAGATTCGCCGGAGCC 1446
```

```
Qy   313 AGGAAAATTCAGATTCGAAACATCCCCTCCACCTGCAGTGGAGGTGTTGATGGACTT    372
          ||:::|||||||||||||  :::||||||||||||
Db    81 CysLysLeuGlnIleArgAsnIleProProHisMetGlnTrpGluValLeuAspGlyLeu  100

Qy   373 TTGGCTCAATATGGACAGTGGAGAATCGGAGAACAGCAAGCAGTCAACACAGACAGAAACCGCC 432
          |||||||::::|||||     :::  ||||||||||||||||||||
Db   101 LeuAlaGlnTyrGlyTyrGlyValGluSerCysGluGlnValAlaAsnThrAspThrGluThrAla 120

Qy   433 GTTGTCAACGTCACATGCAACAAGAGAAGAAGCCATGGAAGCTAAGC    492
          |||||||:||||||||||||||||||||||||
Db   121 ValValAsnValArgTyrGlyAlaLysAspGlnAlaArgGluAlaMetAspLysLeuAsn  140

Qy   493 GGGCACAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCGGATGAGAGGTGAGC    552
          ||||||| :::|||  ||||::::
Db   141 GlyPheLeuMetGluAsnTyrArgLeuLeuLysValSerTyrrIleProArgAspGluThrAlaAla 160

Qy   553 TCCCCTTCGCCCCCTCAGCGAGCCCAGGCGTGG------GACCACTCTTCCGGAGCAA 606
          :::  :::||
Db   161 AlaAspAlaProAlaValGlyValArgArgGlyPheAsnProArgGlyTyrProProArgGln 180

Qy   607 GGGCCACGCCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTTCCGGCTGCGGATCCTG 666
          |||||
Db   181 GlySerProSerLeuGlyAlaArgProLysLeuGlnSerAspValProLeuArgLeuLeu  200

Qy   667 GTCCCCACCCAGTTGTTGGTGCCATCATCCGGAAAGGAGGGCTTGACCATAAAGACATC 726
          ||||||||||
Db   201 ValProThrGlnPheValGlyAlaIleIleGlyValGlyGlyAlaThrIleArgAsnIle  220

Qy   727 ACTAAGCAGACAGCCCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGAGCTCCAGAG 786
          ||||||||||||||||
Db   221 ThrLysGlnThrHisSerLysIleAspIleHisArgLysGluAsnAlaGlyAlaAlaGlyGlu 240

Qy   787 AAGCCTGTCACCATCCATGCCACCCCAGAGGGACTTCTGAAGCATGCCGGATCCGCATGATTCTT 846
          |||||||||||||||||  :::||
Db   241 LysProIleThrValHisSerThrProGluGlyGlyCysSerSerAlaCysArgArgAsnIleMet  260

Qy   847 GAAATCATGCAGAAGAGGCAGATGAGACAACAAACTAGCCGAAGAGATTCCTCTGAAAATC 906
          ||||  :::||||  ||  :::
Db   261 GluIleMetGlnLysGluAlaIleAlaIleAspThrLysIleThrLysIleProLeuLysIle  280

Qy   907 TTGGCACCAATGGCTTGGTTGGAAGACTGATTGGAAGAAGAAGGCAGAAATTTGAAGAAA 966
          ||||:::||||||:||  |||
Db   281 LeuAlaHisAsnAsnPheValGlyArgLeuIleGlyArgLeuGlyLysGluGlyLyrArgAsnLeuLysLys 300

Qy   967 ATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATATAC 1026
          :::||| |||  |||
Db   301 IleGluGlnThrAspThrArgLysIleThrIleSerProLeuGlnAspLeuThrLeuTyr 320

Qy  1027 AACCCGGAAGAGAACCATCACTGTGAAGGGCACAGTTGAGCGCTGTGCAGTGCTGAGATA 1086
          ||||||||||||||| :::||||||
Db   321 AsnProGlnArgThrIleThrValLysGlyThrLeuAsnAlaCysArgArgAlaGlyGlu 340

Qy  1087 GAGATTATGAGAAGCTGCGTGAGGCCTTTGAAAATGATATGCGTGGCTGTTAACCAACAA 1146
          |||||||||||||:::|||||
Db   341 GluIleMetLysLysValArgGluSerTyrGluAsnAspValAlaAlaMetHisLeuGln 360

Qy  1147 GCCAATCTGATCCCAAGGGTTGAACCTCAGCGCCCACTTGGCATCTTTTCAACAGGACTGTCC 1206
          ::||||||||  |||||||||||||||:::||
Db   361 SerAsnLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProGlyAlaAlaSer  380

Qy  1207 GTGCTATCCTCCACAGGACAGGCACTGTCCTCCGCCCCTACCACCCCTTC 1266
          |||||||||||||||||||||||
Db   381 GlyGlyIleSerProSerValValSerGlyTyrProProGlyAla------  395

Qy  1267 ACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTTGCCCGGTTCCCG 1326
          ::|||||||||  ||||||||||
Db   396 ------GlnAlaGlyTyrGlnSer----------------------------------  403

Qy  1327 CATCATCACTCTTATCCAGACAGCAGGATTGTGAATCTCTCATCCCAACTGGAGGTGTG 1386
          |||||||||||||||||
Db   404 ------AlaGlnMetGluSerGluThrValHisLeuPheIleProAlaLeuAlaVal  420
```

```
DB:                              11                    Gaps:           8

Q8C2J9     US-09-270-437D-6 (1-3412) x Q8C2J9 (1-579)

Qy    73 ATGAACAAGCTTTACATCGGAGAACTGGACCCGAGCCCGTCACCGCCAGCGATGTCATCACC 132
                |||||||||||||||||||||||||||||||||||||||| ||||
Db     1 MetAsnLysLeuTyrIleGlyAsnLeuAspProSerProSerProValIleIleSer 20

Qy   133 CTCTTTGGGGACAGGAGGAAGCTGCCCTGGCGGGACAGGAGGCGCCTGCTGAAGTCCGGCTACGCC 192
               :::||| |||||||||||||||||||||||||||||||||:::||||||
Db    21 ValPheLysAspAlaLysIleProValAlaGlyProPheLeuValLysThrGlyTyrAla 40

Qy   193 TTCGTGGACTACCCCGACCAGAACTGGGCCATCCGCGACCTGAGACCCTCCGGGTAAA 252
               ||||| |||||||||||||||||||:::||:|||||||||||||
Db    41 PheValAspCysProAspGluGlyTrpAlaLeuLysAlaIleGluAlaLeuSerGlyLys 60

Qy   253 GTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGTCAAGGAGGC 312
               :::|||||||||||||||||||||||||:||||||||||||||||
Db    61 MetGluLeuHisGlyLysProMetGluValGluHisSerValProLysArgGlnArgIle 80

Qy   313 AGGAAAATCAGATTGAAACATCCCTCCACCTGCAGTGGGAGGTGTTGGATGGACTT 372
               ||| ||||||||||||||||||||||||||||||||||||||||||
Db    81 ArgLysLeuGlnIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100

Qy   373 TTGGCTCAATATGGACAGTGGAGAATGTGGAACAAGTCAACACAGACAGAAACCGCC 432
               :::||||||||||||||||||||:::|||||||||||||||
Db   101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSerGluThrAla 120

Qy   433 GTTGTCAACGTCACATATGCAACAAGAAGAAGAGCAAAAATAGCCATGGAGAGCTAAGC 492
               ||||||||||||||||||||:::||||||||||||:::||
Db   121 ValValAsnValThrTyrSerLysAspGlnAlaArgGlnAlaLeuAspLeuAsnLeuAsn 140

Qy   493 GGGCATCAGTTGAGAACTACTCCTTCAGATTTCCTACATCCCGATGAAGAGGTGAGC 552
               |||:|| |||||||||:::||||:::||:|||||||| ::::
Db   141 GlyPheGlnLeuGluAsnPheThrLeuLysValAlaTyrIleProAspGluThrAlaAla 160

Qy   553 -----TCCCCTTCGCCCCCCAGCGAGCCCAGCTGGGGAC-------CACTCTCC 597
               :::||||||||||||||||||||||||||||
Db   161 GlnGlnAsnProSerProGlnLeuArgGlyArgArgGlyValProGlnGlyGlySerSer 180

Qy   598 CGGGAGCAAGGCCACGCCCCTGGGGCACTCTCAGGCCAGACAGATGATTTCCGGCTG 657
               |||:: :::||||||| ::: |||
Db   181 ArgGlnAla------SerProGlySerValSerLysGlnLysProCysAspLeuProLeu 198

Qy   658 CGGATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGGCTTGACCATA 717
               |||:|||||||||||||:::||||||||:::||||||||||||:::||||||||
Db   199 ArgLeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyValAlaThrIle 218

Qy   718 AAGAACATCACTAGCAGACCCAGTCCCGGTAGATATCCGGTAGAAAGAGAACTCTGGA 777
               :::||||||||||||||||||||||||||||||||||||||||||||
Db   219 ArgAsnIleThrArgGlnThrGlnSerLysIleAspValHisArgLysGluAsnThrGly 238

Qy   778 GCTCGAGAGAGCCTGTCACCATCCATGCACCCCAGAGGGGGACTTCTGAAGCATGCCGC 837
               |||::: |||||||||||||:::||||||||||||||
Db   239 AlaAlaGluGluLysSerIleThrIleLeuSerPheSerAlaAlaCysLys 258

Qy   838 ATGATTCTTGAAATCATGCAGAAGAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCT 897
               ||||||||||||||||||:::|||||||||||||||||||||||
Db   259 SerIleLeuGluIleMetHisLysGluAlaGlnAspIleLysPheThrGluGluPro 278

Qy   898 CTGAAAATCTGGCACACAATGGCTTGGTTGGAGACTGATTGGAAAGAGGCAGAAAT 957
               ||||||||||||||||||||||||||||:::||||||||||||||||
Db   279 LeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuGlyLysGluGlyArgAsn 298

Qy   958 TTGAAGAAATTGAACATGAAACACGGACCAAGATAACAATCCATCTTTCAGGATTTG 1017
               ||||||||:::||||||||||||||||||||||||||||||||
Db   299 LeuLysLysIleLeuGluGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnLeuLeu 318

Qy  1018 AGCATATACAACCCGGAAGAACCATCACTGAAGGGCACAGTTGAGGCCTGTGCCAGT 1077
               :::||||||||||||||||||||||||||||||||||||||:::
Db   319 ThrLeuTyrAsnProGluArgThrIleThrValGlySerValGlyThrCysAlaLys 338

Qy  1078 GCTGAGATAGAGATTATGAGAAGAGCTGCGTGAGGCCTTTGAAAATGATAATGTCGTGTT 1137
```

```
Db   421 GlyAlaAlaIleIleGluGlyLysGlnHisIleLysGlnLeuSerArgPheAlaGlyAla 440

Qy  1447 TCTATCAAGATTGCCCTGCGGAAGGCCAGACCTCAGCGAAAGGATGTCATCATCACC 1506
               ||||||||||||||||||||||| ||| :::|||||||||||||||
Db   441 SerIleLeuSerAlaProAlaAspGlyIleAspLysGlnArgMetValIleIleSer 460

Qy  1507 GGGCCACCGGAAGCCAGTTCAAGGGACGGACGGATCTTTGGGAAACTGAAAGAGAGAA 1566
               :::|||||||||||||||||||||:::|||:::||||
Db   461 GlyProProGluAlaGlnPheLysAlaGlnIleGlyArgIlePheGlyLysLeuLysGluGlu 480

Qy  1567 AACTTCTTTAACCCAAAGAAGAGAGAGAGTGAAGGGCCATATCAGAGTGCCCTCTCC 1626
               |||||||||||||||||||:::|||:::||||||||
Db   481 AsnPhePheGlyLysProLysLysArgGluValLysAlaHisIleLysValProSerPhe 500

Qy  1627 ACAGCTGGCGCGGGTGATTGGCAAGACGTGGCAAGACCGTGAACGAACGTCAGAACTTAACC 1686
               ||||||||||||||||||||||||||||||||||||||
Db   501 AlaAlaGlyAlaGlyValIleGlyLysGlyLysThrValAsnGluLeuGlnLeuAsnLeuThr 520

Qy  1687 AGTGGAGAAGTCATCGTGCGTCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTGATCGTC 1746
               ||||||||||||||||||||||||||||||||||||||||||||||
Db   521 SerArgGluValValValValProArgArgAspGlnThrProAspGluThrAspGlnAspAspSerLeu 540

Qy  1747 AGAATTATCGGGCACTCTTTGCTCTAGCCAGACTCGCACACGCAAGATCAGGGAAATTGTA 1806
               :::|||||| |||||||||||:::|||||||||||||||
Db   541 LysIleThrGlyHisSerPheTyrAlaSerGlnGlnLeuArgGlnLysIleArgGlyIleIleIle 560

Qy  1807 CAACAGGTGAAGCAGCAGCAGCAGAAATACCCT 1839
               |||:::::::||||| |||||||
Db   561 SerGlnValArgArgGlnGlnProLysPro 571
```

```
RESULT 12
Q8C2J9

ID   Q8C2J9        PRELIMINARY;      PRT;     579 AA.
AC   Q8C2J9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Insulin-like growth factor 2.
GN   IGF2BP3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NOD; TISSUE=Thymus;
RX   MEDLINE=22354683; PubMed=12466851;
RA   the FANTOM Consortium;
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573 (2002).
DR   EMBL; AK088465; BAC40370.1; -.
DR   MGD; MGI:1890359; Igf2bp3.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE 579 AA; 63551 MW; 937B601A95D06B77 CRC64;
```

```
Alignment Scores:
Pred. No.:                      2.31e-137      Length:           579
Score:                          1962.50        Matches:          391
Percent Similarity:             79.39%         Conservative:     79
Best Local Similarity:          66.05%         Mismatches:       91
Query Match:                    32.06%         Indels:           31
```

```
Qy    339 AlaAlaGluGluIleMetLysLysIleArgGluSerTyrGluAsnAspIleAlaSerMet 358
              ||||||                      ::::|||||       :::|||
Db   1138 AACCACAAGGCCAATCTGATCCCAGGGTTGAACTCAGGGCACTTGGCATCTTTTCAACA 1197
                                          ||:::|||
Db    359 AsnLeuGlnAlaAsnLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPhe------ 376

Qy   1198 GGACTGTCCGTGCTATCTCCACCAGCAGGGCCCGGAGCTCCCCGCTGCCCCCTAC 1257
              |||||                          |||  |||||
Db    377 --------------------------ProProThrSer------GlyMetProProProThr------ 386

Qy   1258 CACCCCTTCACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTGGC 1317
              ||:||                        |||:::|||   |||
Db    387 --------------SerGlyProProSerThrLeuThrPro------------------ 395

Qy   1318 CCGGTTCCCGCATCATCACTCTTATCCAGACGAGGAGATTGTGAATCTTCATCCCAACC 1377
              ||:::||                      ::|||
Db    396 ProTyrProGlnPheGluGln---SerGlnThrGluThrValHisLeuPheIleProAla 414

Qy   1378 CAGGGTGTGGGCGCCATCATCGGGAAGAAGGGGGCACACTCAAACAGCTGGCGAGATTC 1437
              ||:::||                      |||||||||:::|||||
Db    415 LeuSerValGlyAlaIleIleGlyLysLysGlyGlnHisIleLysLysGlnLeuSerArgPhe 434

Qy   1438 GCCGGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGAGCGTCAGCGAAAGGATGGTC 1497
              ||:::|||                      |||||
Db    435 AlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAlaLysValArgMetVal 454

Qy   1498 ATCATCACCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGACCGATCTTTGGGAAACTG 1557
              |||||                        |||:::||||||
Db    455 IleIleThrGlyProProGluAlaGlnPheLysAlaGlnAspArgIleTyrGlyLysIle 474

Qy   1558 AAAGAGGAAAAACTTCTTTTAACCCCAAAGAAGAAGTGGAAGCTGGAAGTGCATATCAGAGTG 1617
              ||:::||                      |||||||||:::|||
Db    475 LysGluGluAsnPheValSerProLysLysLysTrpLysLeuGluAlaHisIleArgVal 494

Qy   1618 CCCTCTTCCACAGCTGGCCGGGTGATTGGCCAAGGTGGCAAGAACCGTGAACAACTGCAG 1677
              |||||                        |||||||||:::|||||
Db    495 ProSerPheAlaAlaGlyAlaArgValIleGlyLysGlyGlyLysThrValAsnGluLeuGln 514

Qy   1678 AACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAA 1737
              ::||:::||||||                |||||||||:::||||||
Db    515 SerLeuSerSerAlaGluValIleValProArgAspGlnThrProAspGluAsnGluGln 534

Qy   1738 GTGATCGTCAGAATTATCGGGCACTTCTTTGCTCAGCCACAGACTGCACAGCGCAAGATCAGG 1797
              |||:::||                      ::|||   |||:::|||
Db    535 ValValValIleThrGlyHisPheTyrAlaCysHisGlnValAlaGlnArgLysIleGln 554

Qy   1798 GAAATTGTACAACAGGTGAAGCAGCAGGACCAGAAA 1833
              ::||:::||                ::|||
Db    555 GluIleLeuThrGlnValLysGlnHisGlnGln 566
```

RESULT 13
```
ID   Q9CPN8          PRELIMINARY;     PRT;   579 AA.
AC   Q9CPN8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   10 days embryo cDNA, RIKEN full-length enriched library,
DE   clone:2610036B18, full insert sequence (Igf2 mRNA-binding protein 3)
DE   (Insulin-like growth factor 2, binding protein 3).
GN   IGF2BP3 OR 2610101N11RIK OR MIMP3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryo;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
```

```
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Mori H., Sakakibara S., Imai T., Nakamura Y., Iijima T., Suzuki A.,
RA   Yuasa Y., Takeda M., Okano H.;
RT   "Expression of mouse igf2 mRNA-binding protein 3 and its implications
RT   for the developing central nervous system.";
RL   J. Neurosci. Res. 0:0-0(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain, and Olfactory epithelium;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh L.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain, and Olfactory epithelium;
RA   Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AK011689; BAB27779.1; -.
DR   EMBL; AB046173; BAB19755.1; -.
DR   EMBL; AB045138; AAH45138.1; -.
DR   EMBL; BC045138; AAH45138.1; -.
DR   EMBL; BC049082; AAH49082.1; -.
DR   MGD; MGI:1890359; Igf2bp3.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE  579 AA;  63574 MW;  CABD9A4355B392B7 CRC64;
```

Alignment Scores:
Pred. No.:   5.45e-137   Length: 579

Score: 1957.50   Matches: 390
Percent Similarity: 79.39%   Conservative: 80
Best Local Similarity: 65.88%   Mismatches: 91
Query Match: 31.98%   Indels: 31
DB: 11   Gaps: 8

US-09-270-437D-6 (1-3412) x Q9CPN8 (1-579)

```
Qy    73 ATGAACAAGCTTTACATCGGGAACTCGGAACCCCGCCGTCCGCCGTGAGCCCGTCACCGGCAG 132
            ||||||||||||||| ||||||| ||||||||||||||||||
Db     1 MetAsnLysLeuTyrIleGlyAsnLeuSerAspHisAlaGlyProAlaAspLeuGluSer 20

Qy   133 CTCTTGGGGACAGGAGCTGCCCTGGCGCGGACAGGTCCTGCTGAAGTCCGGCTACCGCC 192
            ::||| |||| |||:||||||||| ||||::|||:||||||||||||
Db    21 ValPheLysAspAlaLysIleProValAlaGlyProPheLeuValLysThrGlyTyrArgAla 40

Qy   193 TTCGTGGACTACCCCGACCAGAACTGGGCCATCCGCGCCATCGAGACCCCTCGGGTAAA 252
            ||| |||| ||||||||||||||||||||
Db    41 PheValAspCysProAspGluGlyLyrpAlaLeuLysAlaIleGluAlaLeuSerGlyLys 60

Qy   253 GTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAGCTAAGGAGC 312
            ::||||||||||||||||::||||||:||||:||||||||:||
Db    61 MetGluLeuHisGlyLysProMetGluValGluHisSerValProLysArgGlnArgIle 80

Qy   313 AGGAAAATTCAGATTCGAAACATCCCTCCACCTGCAGTGGGAGGTGTTGATGGGACTT 372
            :::|||||||||||||||||||||:|||||||||||||||||||
Db    81 ArgLysLeuGlnIleArgAsnIleProHisLeuGlnTrpGluValLeuAspSerLeu 100

Qy   373 TTGGCTCAATATGGGACAGTGGGAGAATGTGGAACAAGTCAACACAGAAACCGCC 432
            |||||| :||||:||||||||||||||||||||
Db   101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSerGluThrAla 120

Qy   433 GTTGTCAACGTCACATATGCAACAGAGAAGAGCAAAAATAGCCATGGAGAAGTCAAGC 492
            ||||||||||||||:||||||||||||::||: |||||:||:||
Db   121 ValValAsnValThrTyrSerSerLysAspGlnAlaArgGlnAlaLeuAspAspLysLeuAsn 140

Qy   493 GGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGATGAAGAGTGAGC 552
            |||||||||||||||||||||||||||||||||||||||||
Db   141 GlyPheGlnLeuGluGluAsnPheThrLeuLysValAlaTyrIleProMetProAlaAla 160

Qy   553 ---------TCCCCTGCCCCTCAGCGAGCCCAGCTGGGAC---------CACTCTTCC 597
            :: |||||||||:|||||||||||||| |||
Db   161 GlnGlnAsnProSerProGlnLeuArgGlyArgArgGlyGlyProGlyGlyArgGlySerSer 180

Qy   598 CGGGAGCAAGGCCAGCGCCCCTGGGGGCACTTCTCAGGCCCAGCACAGATTTGATTTCCCGCTG 657
            |||::: :: ||||| ||||||||||||||| :||
Db   181 ArgGlnAla------SerProGlySerValSerLysSerGlnIleLysProCysAspLeu 198

Qy   658 CGGATCCTGGTCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATA 717
            |||||||:||||||||||||||||||||||
Db   199 ArgLeuLeuValProThrGlnPheGlnValAlaIleIleGlyLysGluGlyLysAlaThrIle 218

Qy   718 AAGACATCACTAAGCAACAGACGACCCAGTCCCGGTAGATATCCATAAGAAGAACTCTGA 777
            ::|||||| :||||||||||||:||||:|||||||
Db   219 ArgAsnIleThrLysGlnThrGlnSerArgLysIleAspValHisArgLysSerGluAsnThrGly 238

Qy   778 GCTGCAGAGAAGCCTGTCACCATCCATGCAAGCGGGGACTTCTGAAGCATGCCGC 837
            ||||||||||||| ::||||||||||||||:::::
Db   239 AlaAlaAlaGluLysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCysLys 258

Qy   838 ATGATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTGCCGGAGATTCCT 897
            |||||||||||||||| ||||||||| |||||||||||||||||
Db   259 SerIleLeuGluIleMetHisLysGluAlaGlnAspIleLysPheThrGluIleIlePro 278

Qy   898 CTGAAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAGGCAGAAAT 957
            ||||||||||||||||||||||||||:|||||| |||||||||||
Db   279 LeuLysIleLeuGluLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsn 298

Qy   958 TTGAGAAAATTGAACATGAAACAGGAACAAGATAACAATCTCATCTTTGCAGGATTTG 1017
            ||||||| |||| ||||| ||||||||||||||:||
Db   299 LeuLysLysIleGluGlnLnAspAspThrArgTyrIleSerProLeuGlnGluLeu 318

Qy  1018 AGCATATACACACCCGGAAAGAACAACCATCACTGTGAGGGCACAGTGAGCGTGCCAGT 1077
            ::::::: ::||||||||||||| :::||||||||||||||||| ||||
Db   319 ThrLeuTyrAsnProGluArgThrIleThrValLysGlySerValGluThrCysAlaLys 338

Qy  1078 GCTGAGATAGAGATTATGAAGAGAGCTGGCTGAGGCCTTGAAAATGATGATGTGGCTGTT 1137
            :::::::|||||||:|||::: ::|||||||||| :::: :||| :::::::
Db   339 AlaGluGluIleIleMetLysLysIleArgGluSerTyrGluValAsnAspIleAlaSerMet 358

Qy  1138 AACCAACAAGCCAATCTGATCCCGAAGGTTGAACCTCAGCGACTTGGCATCTTTCAACA 1197
            |||| |||||||| :||||:|||||||||||||||||||||||||
Db   359 AsnLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPhe------ 376

Qy  1198 GGACTGTCCGTGCTATCTGCCGGGACAGGTCCTGCTCCCCCGGAGTCTGCCCCCCTAC 1257
            |||| :::::::::::::::::::::::|:||||||
Db   377 -------------------------GlyMetProProThr-------- 386

Qy  1258 CACCCCTTCACTACCCACTCCGGATACTTCTCCAGCTGTACCCCATCACCAGTTTGGC 1317
            ||||| ::::|||
Db   387 --------------SerGlyProSerThrLeuThrPro----------- 395

Qy  1318 CCGTTCCGCCATCATCATCTCTATCCAGACGACGAGATTGTGAATCTTCATCCACC 1377
            |||::|||||||:|||||||| |||||||||||||
Db   396 ProTyrProGlnPheGluGln---SerGluThrGluThrValHisLeuPheIleProAla 414

Qy  1378 CAGGCTGTTGGGCGCCATCATCGGGAAGAAGGGGCACACATCAAACAGCTGGGAGATTC 1437
            :::||||||||||||||:::||||||||||||||||||
Db   415 LeuSerValGlyAlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgPhe 434

Qy  1438 GCCGGAGCCTCATCAAGATTGCCCCTGCGAAGGCCCAGACGTCAGCGAAGGATGGTC 1497
            :::||||||||||||:||||||||||:|||||||||||||||||||
Db   435 AlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLysValArgMetVal 454

Qy  1498 ATCATCACCGGCGAAGCCCAGTCAAGGACGCCGGATCTTTGGAAACTG 1557
            ||||||||||||||||||||::::||||||||||||||||||||
Db   455 IleIleThrGlyProProGluAlaGlnPheLysAlaGlnIleGlyArgIleTyrGlyLysIle 474

Qy  1558 AAAGAGGAAAACTTCTTTAACCCCAAAGAAGAGTGAAGCTGGAAGGCCATATCAGAGTG 1617
            ||||||| :::|||||||||||||||||||||||
Db   475 LysGluGluAsnPheValSerProLysGluAlaValLysLysLeuGluAlaHisIleArgVal 494

Qy  1618 CCCTCTTCCACAGCTGGCGCGGGTGATTGGCAAGGTGTGAACGAACTGCAG 1677
            :::|||||:|||||||||||||| ::::||| :::::|||
Db   495 ProSerPheAlaAlaGlyArgValIleGlyLysGlyLysThrValAsnGluLeuGln 514

Qy  1678 AACTTAACCAGTGCAGAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAATGAGGAA 1737
            :::||||||:::::||| ||||||||||:|||||||||||||||||
Db   515 SerLeuSerSerAlaGluValValProArgAspGlnThrProAspGluAsnAspGln 534

Qy  1738 GTGATCGTCAGAATTATCGGGCACTTCTTTGCTGTCCAGACTGCCACAGCCCAAGATCAGG 1797
            ||||||:||||||||||||||||||:||||||||||||||||||||||||
Db   535 ValValValLysIleThrGlyHisPheTyrAlaCysGlnValAlaGlnArgLysIleGln 554

Qy  1798 GAAATTGTACAACAGGTGAAGCAGCAGGAGCAGAAA 1833
            ||||||||:|||||||||||||||||||| :::|||::::
Db   555 GluIleLeuThrGlnValLysGlnHisGlnGlnGln 566
```

RESULT 14
O00425
ID   O00425   PRELIMINARY;   PRT;   579 AA.
AC   O00425;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative RNA binding protein KOC (KOC).
GN   KOC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mueller-Pillasch F., Lacher U., Wallrapp C., Et AL.;
RL   Oncogene 0:0-0(0).
RN   [2]

```
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RA   Mueller-Pillasch F., Lacher U., Wallrapp C.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; U97188; AAC35208.1; -.
DR   EMBL; U76705; AAD09223.1; -.
DR   GO; GO:0005737; C:cytoplasm; TAS.
DR   GO; GO:0003723; F:RNA binding; TAS.
DR   GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR   GO; GO:0006412; P:protein biosynthesis; TAS.
DR   GO; GO:0006396; P:RNA processing; TAS.
DR   InterPro; IPR004087; KH dom.
DR   InterPro; IPR004088; KH type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
SQ   SEQUENCE   579 AA;  63720 MW;  AE5C3AEE3C135C5 CRC64;
```

```
Alignment Scores:
Pred. No.:                            1.81e-136    Length:        579
Score:                                1950.50      Matches:       386
Percent Similarity:                   79.02%       Conservative:  81
Best Local Similarity:                65.31%       Mismatches:    95
Query Match:                          31.87%       Indels:        29
DB:                                   4            Gaps:          8
```

```
US-09-270-437D-6 (1-3412) x O00425 (1-579)

Qy    73 ATGAACAAGCTTTACATCGGGAACCTGAGCCCCGTCACCGGCGACGACCTCCGGCAG 132
          |||||||
Db     1 MetAsnLysLeuTyrIleGlyGlyAsnLeuSerGlyAsnAlaAlaProSerAspLeuGluSer 20

Qy   133 CTCTTTGGGGACAGGAAGCTGCCCCGGGGACAGTTCCTGCTGAGTCCGGCTACGCC 192
          :::||    |||    ||||||||       ||||||||||    ::||||||||
Db    21 IlePheLysAspAlaLysIleGlyProValSerGlyProPheLeuValLysThrGlyTyrAla 40

Qy   193 TTCGTGGACTACCCCGACCAGAACTGGGCCATCGGCCATGGGCCATCCTCGGGTAAA 252
          ||||        ||||||||||||||     |||||||||||||||||||||||||
Db    41 PheValAspCysProAspGluSerTrpAlaLeuLysAlaIleGluAlaLeuSerGlyLys 60

Qy   253 GTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCCTAAAAAGCTAAGGAGC 312
          :::||||||    ||||||||||||||    ||||||||    :||:||    ||||||
Db    61 IleGluLeuHisGlyLysProIleGluValAspTyrSerValProLysArgGlnArgIle 80

Qy   313 AGGAAAAATTCAGATTCGAAACATCCCTCACCTGACGTGGAGGTGTTGATGGACTT 372
          |||||    :::||||||||||||||||    |||||    ||||||||||||
Db    81 ArgLysLeuGlnIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100

Qy   373 TTGGCTCAATATGGACAGTGGAGAATGGAGAACAAGTCAACACAGACAGAAACCGCC 432
          ||||        ||||||||||:::    ||||||||||||||||||||||||
Db   101 LeuValGlnTyrGlyValValGluSerGluGlnValAsnThrAspSerGluThrAla 120

Qy   433 GTTGTCAACGTCACATATGCACAAGAAGAAGACAAAAATAGCCATGGAGAAGCTAAGC 492
          |||:::||:||    ||||||||||||    ||||||||||    ||||||||||:::
Db   121 ValValAsnValThrTyrSerLysLysAspPheLeuLysValAlaTyrIleProAsnLeuAsn 140

Qy   493 GGGCATCAGTTTGAGAACTACTCCTCAAGATTCCTACATCCCGGATGAAGAGGTGAGC 552
          ||||        :::||||||||    ||||||||||||||    --------GGGGACCACTCCTTCCCGG
Db   141 GlyPheGlnLeuGluAsnPheThrLeuLysValAlaTyrIleProAspGluGluMetAlaAla 160

Qy   553 TCCCCTTCGCCCCCCTCAG------CGAGCCCAGCGT-----GGGGACCACTCCTTCCCGG 600
          ||||    :::|||    ||    ||||||||||:::
Db   161 GlnGlnLeuAsnProLeuGlnGlnProArgGlyLysArgGlyLeuGlnArgGlyLySerSer 180

Qy   601 GAGCAAGGCCACCCCAGTTCCTCAGGCCAGACAGATTGATTTCCGCTGCGG 660
          ||||||||        ::|||||    ||    ::    ||
Db   181 ArgGlnGly---SerProGlySerValSerLysLysGlnLysProCysAspLeuProLeuArg 199
```

```
Qy   661 ATCCTGGTCCCCACCCCAGTTTGTTGGTGCCATCATCGGGAAAGGAGGGCTTGACCATAAAG 720
          :::||||||||||||||||||||||    ||    ::|||||    |||    :::||
Db   200 LeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGlyGlyAlaThrIleGluArg 219

Qy   721 AACATCACTAAGCAGACGACCCGGTAGATATCATAGAAAAGAGAACTCTGGAGCT 780
          :::||||||||||||||||    |||||||||||||    ::||:||||||||||
Db   220 AsnIleThrLysGlnThrGlnSerLysIleGluAspValHisArgLysGluAsnAlaGlyAla 239

Qy   781 GCAGAGAAGCCTGTCACCATCCATGCCACCCGAGGGGGACTTCTGAAGCATGCCGCATG 840
          :::||||||    ||||||||||||||||||||    |||||    ||||||:::
Db   240 AlaGluLysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCysLysSerSer 259

Qy   841 ATTCTTGAAATCATGCAGAAAGAGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG 900
          |||||||||||    ||||||||||    ::||    ||||||||||||||||||
Db   260 IleLeuGluIleMetHisLysGluAlaGlnAspIleLysPheThrGluGlnIleProLeu 279

Qy   901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGGTTGGAAAGAAGGCAGAAAATTG 960
          ||||||||||    ::||||||||    ||||||    :::||||||||||
Db   280 LysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLyGluGlyArgAsnLeu 299

Qy   961 AAGAAAATTGAACATGAAACAGGACCAAGATAACAATCCTCATCTTTGCAGGATTTGAGC 1020
          |||    :::||    ||||||||||||||||||||||||||    :::|||
Db   300 LysLysIleGluGluAspThrArgAspThrArgIleThrIleSerProLeuGlnGluLeuThr 319

Qy  1021 ATATACACCCGGAAAGAACCAATCTGATCCCAGGGTTGAACCTCAGCCACTTGGACATCTTTTCAACAGGA 1200
          :::||||||||    ||||||||||||    ||||||||||    ||    |||||||||
Db   320 LeuTyrAsnProGluArgThrIleThrValLysGlyAsnValGluThrCysAlaLysAla 339

Qy  1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGGCTTTGAAAAATGATATGTGGCTGTTAAC 1140
          |||    ||||||||||||||||||||||    :::||||||    :::||
Db   340 GluGluGluMetLysLysIleGluArgGluSerTyrGluAsnAspIleAlaSerMetAsn 359

Qy  1141 CAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCCACTTGGACATCTTTTCAACAGGA 1200
          |||||||
Db   360 LeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLyLeuPhe-------- 376

Qy  1201 CTTGTCCGTGTATCTCCACCAGGCAGGGCCCCGGAGCTCCCCGGGGACTCCCCCTACCAC 1260
          |||||
Db   377 --------------------ProProThrSer-----GlyMetProProProThr----- 386

Qy  1261 CCCTTCACTACCCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTTGGCCCG 1320
          ||||||
Db   387 -------------SerGlyProProSerAlaMetThrPro------------Pro 396

Qy  1321 TTCCCCATCATCACTCCTTATCCAGGACGAGGAGAGTTGAATCTCTCATCCCAACCCAG 1380
          :::||||||
Db   397 TyrProGlnPheGluGln---SerGluThrValHisGlnPheLeuProAlaLeu 415

Qy  1381 GCTGTGGGCGCCATCATCGGGAAGAAGAGGGGCCACATCAAACAGCTGGCGAGATTCGCC 1440
          :::||||||||||||||||||||||||||||||||||||||||    ::|||||||
Db   416 SerValGlyAlaIleIleGlyLysLysGlyGlnHisIleLysGlnLeuSerGlnArgPheAla 435

Qy  1441 GGAGCCTCATCCAAGATTGCCCTGCCGAAGGCCCAGAAGCTCAGCGAAGGATGGTCATC 1500
          ||||||||||    |||||||||||||||||||||||||||||    :::||
Db   436 GlyAlaSerIleLeuGlyAlaProAlaLysAlaProAlaLysAlaLysArgMetValIle 455

Qy  1501 ATCACCGGGCCACCGAAGCCCAGTTCAAGCCCCAGGGACGGGCGGATCTTTGGGAAACTGAAA 1560
          :::||||||||||||||||||||||||||||||||    ||    ||||||
Db   456 IleThrGlyProProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysLeuLys 475

Qy  1561 GAGGAAAACTTCTTTAACCCAAAGAAGAAGTGCAAAGTGGCAAGGCGCATATCAGATGTGCCC 1620
          :::||||||||||||||    |||||||||||||||||||||    :::||
Db   476 GluGluAsnPheValSerProLysProLysValAlaProGlyLysLysAlaHisIleArgValPro 495

Qy  1621 TCTTCCAGCTGCGGGCGTGGCGGGGTGATTGGCTGGGTGGCAAAGTGGCGCAAGTGCCAAGAAC 1680
          |||                     |||||||||||||||||||||||||||||||:::
Db   496 SerPheAlaGlyAlaGlyArgValIleGlyLyLysGlyAlaThrValLysGluLysGlnAsn 515

Qy  1681 TTAACCAGTGCAGAAGTCATCGTGCCTCAGGCCAGACAGATTGATTTCCGGCTGCGG 1740
          :::||||||    :::|||||||||||    :::|||||    ||:::||
Db   516 LeuSerSerAlaGluValValProArgGlyGlnThrProAspGluAlaAspAspGlnVal 535
```

```
Qy   1741  ATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCACAGCGCAAGATCAGGGAA 1800
                 :::||||||   ||||||||  :::||||   ||| |||||||||   :::|||
Db    536  ValValLysIleThrGlyHisPheTyrAlaCysPheArgGlnValAlaGlnArgLysIleGlnGlu 555

Qy   1801  ATTGTACACAGGTGAAGCAGAGACGAGGAGCAGAAA 1833
                 |||||||| :::||||||||
Db    556  IleLeuThrGlnValLysGlnHisGlnGln 566

RESULT 15
Q8IGK4
ID   Q8IGK4       PRELIMINARY;       PRT;      580 AA.
AC   Q8IGK4;
DT   01-MAR-2003  (TrEMBLrel. 23, Created)
DT   01-MAR-2003  (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003  (TrEMBLrel. 25, Last annotation update)
DE   RE72930p.
GN   IMP OR CG1691.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Y;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BT001739; AAN71494.1; -.
DR   FlyBase; FBgn0030235; Imp.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 4.
DR   SMART; SM00322; KH; 4.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
SQ   SEQUENCE  580 AA;  63407 MW;  86EA1F5D64523C02 CRC64;

Alignment Scores:
Pred. No.:                          580
Score:           3.09-56            Length:               580
                 873.50             Matches:              231
Percent Similarity:  51.30%         Conservative:         104
Best Local Similarity:  35.38%      Mismatches:           187
Query Match:         14.27%         Indels:               131
DB:                                 Gaps:                  18

US-09-270-437D-6 (1-3412) x Q8IGK4 (1-580)
```

```
Db    71  AsnProTyr-------                                                 ProGlyMetProGlyProGlyArgGln 82
              ::|                                                         ||||||
Qy   283  GATTACTCAGTCTCTAAAAGCTAGGAGGCAGGAGAAATTCAGATTCGAAACATCCTCCT 342
              ||||||||||   :::||||||||||     ||| |||
Db    10  AspLeuGluLeuSerLysGluAspArgGlu-------GlnIle------PheAspPro 24

Qy   643  ATTGATTTCCGCTGCGGATCCTGGTCCCACCCAGTTGTTGTGTGCCATCATCGGGAAAG 702
              ::|   |||||||||||     :::||| |||
Db    83  AlaAspPheProLeuArgIleLeuValGlnSerGluMetValGlyAlaIleIleGlyArg 102

Qy   343  CACCTGCAGTGGGAGGTGTTGGATGGACTTTGGCTCAATATGGGACGTGGAGAATGTG 402
              |||   |||||||||||    |||:::|||
Db    25  ProLeuAspArgGlnLeuGluGly--------------------------------- 33

Qy   703  GAGGGCTTGACCATAAGAACTCACTAGCAGAGAACATCCCGGGTAGATATCCATAGA 762
              :::  ||||||   |||||   ||||||||| ||||| ::::|||
Db   103  GlnGlySerThrIleArgThrIleThrGlnSerArgAlaArgValAspValHisArg 122

Qy   403  GAACAGTCAACAACAGACACAGAAACCGGTCGTTGCAACGTCACATATGCAACAAGAGAA 462
              :::||   |||||||   :::|||||||  :::
Db    34  --------AlaGlyThrSerArgAlaAlaValGlyLeuAsnGlyValGluPheGlu 49

Qy   763  AAAGAGAACTCTGAGCTGCGAGAGAAGCCTGTCACCATCCATGCCACCCAGAGGGGACT 822
              ||||||||||    |||||:::   |||||||
Db   123  LysGluAsnValGlySerLeuGluLysSerIleThrIleTyrGlyAsnProGluAsnCys 142

Qy   463  GAAGCAAAAATAGCCATGGAAGCTAAGGGGCATCAGTTGAGAACTACTCCTTCAAG 522
              :::||||| :::|||  ||||||   :::|||
Db    50  GlySerLysHisAlaHisLeuAspGlnLeuAspProGlyInTrpLysAlaGlnArgArgSer 65

Qy   823  TCTGAAGCATGCCCATGATTCTTGAAATCATGCAGAAGAGGCAGATGAGACCAAACTA 882
              :::  ||||||   |||:::||||||   :::  |||
Db   143  ThrAsnAlaCysLysArgIleLeuGluValMetGlnGlnAlaIleSerThrAsnLys 162

Qy   523  ATTTCCTACATCCCGATGAAGAGGTGAGCTCCCCTTCGCCCCCTCAGCGAGCCCAGCGT 582
              :::   |||||||||||||||||||||     :::|||
Db    66  -----------------------------------GlnArgAsnGlnArg 70

Qy   883  GCCGAA---------                            GAGATTCCTGAAAATTGGCACCAATGGCTTG 924
              ||||||                                    |||||||||
Db   163  GlyGluLeuSerProGluCysSerGluIleCysLeuLysIleLeuAlaHisAsnAsnLeu 182

Qy   583  GGGGACCACTCTTCCCGGGAGCAAGGCCCACGCCCCCTGGGGCGCCCCCTGGGGGAGCACTTCTCAGGCCAGACAG 642
              ::::
Qy   925  GTTGGAAGACTGATTGGAAAAGAGAGGCAGAAATTGAAGAAATTGAACATGAAACAGGG 984
              ::||||||   |||:::  ||||
Db   183  IleGlyArgIleIleGlyLysSerGlyAsnThrIleLysArgIleMetGlnAspThrAsp 202

Qy   985  ACCAAGATAACAATCCATCTTGCAGGATTTGAGCATTATACAACCCGGAAAGAACCATC 1044
              ||||||    ||||| |||:::  |||
Db   203  ThrLysIleThrValSerSerIleAsnAspIleAsnSerPheAsnLeuGluArgIleIle 222

Qy  1045  ACTGTGAAGGGCCACAGTGAGGGCTGTGCCAGTGCGATGAGATTATGGAAGAAGCTG 1104
              |||||||   :::|||   |||||||
Db   223  ThrValLysGlyLeuIleGluAsnMetSerArgAlaGlyAsnGlnIleSerThrLysLeu 242

Qy  1105  CGTGAGGGCCTTTGAAAATGATATGCTGGCTGTGTTAACCAACAAGCCAATCTGATCCCAGG 1164
              |||:::::::||||||||  ||| :::|||
Db   243  ArgGlnSerTyrGluAsnAspLeuAsnAspIleAsnSerLeuMetPheProGly 262

Qy  1165  TTGAACCTGCAGCGCACTT--------                        GGCATCTTTTCAACAGGACTG 1203
              ||||||||||||||                                    :::|||
Db   263  LeuIleProMetAlaMetSerThrProGlyAsnGlyMetValPheAsnThrSerMet 282

Qy  1204  TCCGTGCTATCTCCACCGCAGCAGGCGCGGAGCTCCCCGTGCCCCCTACCACCCC 1263
              |||||||||||   :::||| |||||||
Db   283  ------------------------------------------Pro 283

Qy  1264  TTCACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTGGCCCG-- 1320
              |||:::|||   |||||||
Db   284  PheProSerCysGlnSerPheAlaMetSerLysThrProAlaSerValValProProVal 303

Qy  1321  TTCCCGCATCATCACTCTTATCCAGACAGGAGATTGAATCTTCATCCCAACCCAG 1380
              |||   :::|||||||||||||||||||
Db   304  PheProAsn-------                            AspLeuGlnLeuThrPheTyrLeuTyrIleProAsnAsn 319

Qy  1381  GCTGTGGGGCCCATCATCGGGAAGAAGGGGCACACATCAAACAGCTGCGAGATTCGCC 1440
              :::|||   |||||||:::|||
Db   320  AlaValGlyAlaIleIleGlyArgGlySerHisIleArgSerIleMetArgPheSer 339

Qy  1441  GGGAAGGCCTTATCAAGATTGCCCCT------                    GACGTCAGCGAA 1488
              |||||||||||||||||||||                              :::|||
Db   340  AsnAlaSerLeuLysIleLeuAlaProLeuAlaAspLysProLeuAspGlnThrGlu 359

Qy  1489  AGGATGGTCATCATCACCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGGACCGGATCTTT 1548
              :::|||||||||||||||||
Db   360  ArgLysValThrIleValGlyThrProGlyGlnInTrpLysAlaGlnInTyrMetIlePhe 379

Qy  1549  GGGAAACTGAAAGAGAGAAACTTCTTTAACCCCAAAGAAGAGAAGAGTGAAGCGGAAGCGCAT 1608
              ||||||||||   :::||||||||
Db   380  GlyLysMetArgGluGluGluGlyPheMetCysGlyThrAspValArgLeuThrValGlu 399

Qy  1609  ATCAGAGTGCCCCTCTTCCACAGTGTGGCCGGGTGATTGGCAAAGGTGGCAAAGACCGTGAAC 1668
              :::
```

```
Db   400  LeuLeuValAlaSerSerGlnValGlyArgIleIleGlyGlyLysGlyGlyGlnAsnValArg  419

Qy  1669  GAACTGCAGAACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAA---ACGCCAGAT 1725
          ||||||||           :::  :::  ::: ||||||       ::||||
Db   420  GluLeuGlnArgValThrGlyValIleLysValIleLeuProGluHisAlaLeuAlaProPro  439

Qy  1726  GAA------AATGAGGAAGTGATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACT 1779
          :::|||||           ||  ::::|||||  |||:::::::||:
Db   440  SerGlyGlyAspGluGluThrProValHisIleIleGlyLeuPheTyrSerValGlnSer  459

Qy  1780  GCACAGCGCCAAGATCAGGGAAAATTGTA----------------------------- 1806
          ||||||||||||||||||      :::::::
Db   460  AlaGlnArgArgIleArgAlaMetMetLeuSerThrAsnProProIleThrLysLys  479

Qy  1807  ----------CAACAGGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCC 1851
                    ::||||||:|||||||||||:::|||           ::||||:::
Db   480  GlnLysAlaAlaLysGluGluGlnGlnGlnSerLeu--AlaGlyAlaAlaS  499

Qy  1852  TCACAGGCGCAGTGAGCTCCCACAGGCACCACCAGCAAAACAACGATGAATGTAGCCC 1911
          ||   ::: |||:||||||       ::||||||:::
Db   499  erSerGlySerGlnGlnGlnGlnProGlnSerProSerGlnGlnGln-------AlaL  516

Qy  1912  TTCCAACACCTGACAGAATGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAGA 1971
          |||||             :::::::::
Db   516  euProProGlnLeuHisHisGlnProValSerSerAlaSerSerSerThrPro----  534

Qy  1972  CCATCTGAGGAATGAGAAGTCTGCGGAGGCGGCCAGGGACTCTGCCGAGGCCCTGAGAAC 2031
                                                       :::
Db   535  ----------------------------------------ProAlaHisHisGlnG  540

Qy  2032  CCCAGGGGGCCGGAGGGGGCCGGGGAAGGTCAGCCCAGGTTTGCCAGAACCACCGAGCCCCG 2091
          |||                                   :::|||  ||||
Db   540  lnGlnAlaSerThrAlaAlaThrSerHisGlnLeuGlnLnGlnGlnProSerProPro-  559

Qy  2092  CCTCCCGCCCCCCAGGGCTTCTGCAGGCTTCAGCC 2126
          |||||||||     ||:::|||
Db   560  -------ProProGlyAsnAlaThrAlaAlaAla  568
```

Search completed: July 16, 2004, 10:44:26
Job time : 203.5 secs

OM nucleic - nucleic search, using sw model

```
               GenCore version 5.1.6
         Copyright (c) 1993 - 2004 Compugen Ltd.
```

Run on: July 20, 2004, 21:16:56 ; Search time 8261 Seconds
                                         (without alignments)
                              12333.829 Million cell updates/sec

Title: US-09-270-437D-6
Perfect score: 3412
Sequence: 1 ggcagcggaggaggcgagga........aaccctgaaaatgtttattt 3412
Scoring table: IDENTITY_NUC
                  Gapop 10.0 , Gapext 1.0
Searched: 27513289 seqs, 1493109276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   EST:*
            1:   em_estba:*
            2:   em_esthum:*
            3:   em_estin:*
            4:   em_estmu:*
            5:   em_estov:*
            6:   em_estpl:*
            7:   em_estro:*
            8:   em_htc:*
            9:   gb_est1:*
           10:   gb_est2:*
           11:   gb_htc:*
           12:   gb_est3:*
           13:   gb_est4:*
           14:   gb_est5:*
           15:   em_estfun:*
           16:   em_estom:*
           17:   em_gss_hum:*
           18:   em_gss_inv:*
           19:   em_gss_pln:*
           20:   em_gss_vrt:*
           21:   em_gss_mam:*
           22:   em_gss_mus:*
           23:   em_gss_pro:*
           24:   em_gss_rod:*
           25:   em_gss_phg:*
           26:   em_gss_vrl:*
           27:   gb_gssI:*
           28:   gb_gss2:*
           29:   gb_gss2:*

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2163	63.4	3592	11	AK049196	AK049196 Mus muscu
2	2006.2	58.8	3262	11	AK077404	AK077404 Mus muscu
3	1862.8	54.6	2964	11	AK044984	AK044984 Mus muscu
4	842.2	24.7	874	13	BQ691878	BQ691878 AGENCOURT
5	841	24.6	905	12	BM453327	BM453327 AGENCOURT
6	830	24.3	851	13	BQ221568	BQ221568 AGENCOURT
7	829.8	24.3	999	12	BM561242	BM561242 AGENCOURT
8	809	23.7	832	13	BX644668	BX644668 DKFZp781G
9	803.4	23.5	1089	12	BM806132	BM806132 AGENCOURT
10	799.2	23.4	836	13	BQ212500	BQ212500 AGENCOURT
11	759.8	22.3	894	13	BU156245	BU156245 AGENCOURT
12	752.6	22.1	781	12	BG575889	BG575889 602598315
13	751.4	22.0	940	14	CD243351	CD243351 AGENCOURT
14	749.8	22.0	824	14	CF593505	CF593505 AGENCOURT
15	746	21.9	889	12	BG748346	BG748346 602705902
16	718.8	21.1	762	13	BU601634	BU601634 AGENCOURT
17	708.6	20.8	4100	11	AK088465	AK088465 Mus muscu
18	707	20.7	2202	11	AK011689	AK011689 Mus muscu
19	698.2	20.5	2780	11	AK013940	AK013940 Mus muscu
20	695	20.4	2718	11	AK044850	AK044850 Mus muscu
21	691.8	20.3	939	13	BQ961314	BQ961314 AGENCOURT
22	689	20.2	724	12	BM977219	BM977219 UI-CF-DU1
23	686.4	20.1	891	13	BU152527	BU152527 AGENCOURT
24	677.8	19.9	918	13	BQ212353	BQ212353 AGENCOURT
25	652.6	19.1	815	13	BI688088	BI688088 603114678
26	634	18.6	647	13	BU674692	BU674692 UI-CF-DU0
27	631.6	18.5	729	10	BE278075	BE278075 601179604
28	625.4	18.3	859	13	BQ890899	BQ890899 AGENCOURT
29	606.2	17.8	759	10	BE622021	BE622021 601440673
30	605.4	17.7	1024	9	AL549055	AL549055 AL549055
31	599.6	17.6	821	14	CA511280	CA511280 UI-R-FJ0-
32	595	17.4	628	12	BI862502	BI862502 603390625
33	594	17.4	607	12	BM724394	BM724394 UI-E-E01-
34	591.8	17.3	759	14	CA320888	CA320888 UI-M-FW0-
35	587	17.2	612	13	BU679082	BU679082 UI-CF-DU1
36	584	17.1	584	13	BX642777	BX642777 DKFZp781I
c 37	571.4	16.7	616	14	CA430674	CA430674 UI-H-FL1-
38	565	16.6	638	12	BM843467	BM843467 K-EST0121
39	561.2	16.4	699	13	BU705477	BU705477 UI-M-FO0-
c 40	553.8	16.2	1078	12	BM462051	BM462051 AGENCOURT
41	548	16.1	594	12	BI858380	BI858380 603383965
42	547.4	16.0	1212	13	BU902212	BU902212 AGENCOURT
43	546.8	16.0	668	13	BY734118	BY734118 BY734118
44	542.6	15.9	845	13	BU155790	BU155790 AGENCOURT
45	541.6	15.9	707	13	BY720350	BY720350 BY720350

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1
AK049196
LOCUS   AK049196   3592 bp   mRNA   linear   HTC 20-SEP-2003
DEFINITION Mus musculus ES cells cDNA, RIKEN full-length enriched library,
            clone:C330012H03 product:HEPATOCELLULAR CARCINOMA AUTOANTIGEN
            homolog [Homo sapiens], full insert sequence.
ACCESSION   AK049196
VERSION   AK049196.1 GI:26093366
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS   Carninci,P. and Hayashizaki,Y.
TITLE   High-efficiency full-length cDNA cloning
JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
MEDLINE   99279253
PUBMED   10349636
REFERENCE   2
AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE   20499374
PUBMED   11042159
```

REFERENCE   3
AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
             Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
             Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
             Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
             Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
             Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
             Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
             sequencing pipeline with 384 multicapillary sequencer
JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE   20530913
PUBMED   11076861
REFERENCE   4
AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
             FANTOM Consortium.
TITLE   Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
             Group Phase I & II Team.
TITLE   Analysis of the mouse transcriptome based on functional annotation
             of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 3592)
AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
             Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
             Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
             Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
             Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
             Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
             Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
             Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
             Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
             Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
             Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
             Muramatsu,M. and Hayashizaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
             Physical and Chemical Research (RIKEN), Laboratory for Genome
             Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
             RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
             Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
             URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
             Fax:81-45-503-9216)
COMMENT   cDNA library was prepared and sequenced in Mouse Genome
             Encyclopedia Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in RIKEN.
             Division of Experimental Animal Research in Riken contributed to
             prepare mouse tissues.
             Please visit our web site for further details.
             URL:http://genome.gsc.riken.go.jp/
             URL:http://fantom.gsc.riken.go.jp/.
FEATURES   Location/Qualifiers
     source   1..3592
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:C330012H03"
                     /db_xref="MGI:2416441"
                     /db_xref="taxon:10090"
                     /clone="C330012H03"
                     /cell_type="ES cells"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
     misc_feature   34..1811
                     /note="HEPATOCELLULAR CARCINOMA AUTOANTIGEN homolog [Homo
                     sapiens] (SPTR|Q9Y6M1, evidence: FASTY, 88.3%ID,
                     100%length, match=1775)
                     putative"
ORIGIN

Query Match   63.4%;   Score 2163;   DB 11; Length 3592;
Best Local Similarity 82.9%; Pred. No. 0;

Matches 2831; Conservative   0; Mismatches 486; Indels   97; Gaps 28;

Qy   38 GGGGGAGCCGCGGGCTCTCGGGGAAGAGACGGATGAGCAAGCTTTACATCGGGAACC   97
          |||||||||||||||||| |||||| ||||||||||||||||||| ||||||||||
Db   3 GGGGGAGCCGCGGGCTCTC-GGGAAGACGCGGATGATGAACAAGCTGTACATTGGGAACC   61

Qy   98 TTGAGCCCCGCCGTCACCGCGCAGACCTTCGGCAGCTCTTCGGGACAGGAAGCTGCCCC   157
          ||||||||||| ||||||||| |||||||||||||||||||||||||||||||||||||
Db   62 TGAGCCCCGCCGTCACCGCGCAGACCTCCGGCAGCTCTTCGGGACAGGAAGCTGCCCC   121

Qy   158 TGGCGGGACAGGTCCTGCTGAGTCCGGCTACGCCTTCGTCGGACTACCCCGACCAGAACT   217
          |||||||||||||||||| || |||||||||||||||||||||||||||| |||||||||
Db   122 TGGCGGGACAGGTCCTACTCAGTCCGGCTACGCCCTTCGTCGGACTACCCCGACCAGAACT   181

Qy   218 GGGCCATCCGCCGCCATCGAGACCCTCGGGTAAGTGGAATTGCATGGGAAAATCATGG   277
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   182 GGGCCATCCGCCGCCATCGAGACCCTCGGGTAAGTGGAATTGCATGGGAAAATCATGG   241

Qy   278 AAGTTGATTACTTCAGTCTCTAAAAAGCTAAGGAGCAGGAAAATTCAGATTCGAAACATCC   337
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||| ||||||
Db   242 AAGTTGACTACTTCAGTCTCTAAAAAGCTAAGGAGCAGAAGAATCCAGATTCGGAACATCC   301

Qy   338 CTCCTCACCTGCAGTGGGAGGTGTTGGATGGACTTTGGCTCAATATGGACAGTGGAGA   397
          || |||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   302 CGCCTCACCTGCAGTGGGAGGTGTTGGATGGGCTGTTGGCTGAATATGGACAGTGGAGA   361

Qy   398 ATGTGGAACAAGTCAACACAGACAGAAACCGCGTTGTCAACGTCACATATGCAACAA   457
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   362 ACGTGGAGCAAGTCAACACAGATACAGAAACTGCCGGTGTCAACGTCACCTATATGACAA   421

Qy   458 GAGAGAAGAGCAAAAATAGCCATGGAGGAGCTAAGCGGGCATCAGTTGAGAACTACTCCTT   517
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   422 GAGAGAGAGGCAAAGCTAGCTATTGAGAAGCTCAGTGGGCATCAGTTTGAGGACTACTCCT   481

Qy   518 TCAAGATTTCTACATCCGGATGAAGAGAGTGAGCTCCCCTTCGCCCCCTCAGCGGAGCC   577
          |||||||||| |||||||||||||||||||||||||||| ||||||||||||||||
Db   482 TCAAGATTTCCTACATCCCCATGAAGAGAGGTGAGCTCTCCTTCACCCCCCTCATCG-----   536

Qy   578 AGCGTGGGGAGCCACCTCTTCCCGGAGCCAAGGCCACGCCACGCCCCTGGGGGCACTTCTCAGGCCA   637
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   537 --------------TGCCCGGGAACAAGGCCCACGCCACGCCCCGGGAGCTCTTCTCAGGCCA   580

Qy   638 GACAGATTGATTTCCCGCTGCGGATCCTGGTTCCCACCCAGTTTGTTGGTGCCATCATCG   697
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   581 GACAGATTGATTTCCCGCTGCGGATCCTGGTTCCCTGTCCCCAGTTTGTTGGTGCCATCATCG   640

Qy   698 GAAAGGAGGGGCTTGACCATAAAGAACATCACTAAGCAGACCCCAGTCCCGGGTAGATATCC   757
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   641 GAAAGGAGGGGCTTGACCATAAAGAACATCACTAAGCAGACCCCAGTCCCGGGTAGACATCC   700

Qy   758 ATAGAAAAGAGAACTGGAGCTGGAGCTGCGAGAGAGCCGTCACCATGCCACCCCAGGAGG   817
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   701 ACAGAAAGGAGAACTGGGGCTGGAGCTGCGAGAGAGCCGTCACAATCCATGCGTACCCCCAGGAAG   760

Qy   818 GGACTTCTGAAGCATGCGCATGATTCTTGAAAATCATGCAGAAGAGGCAGATGAGACCA   877
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   761 GGACATCTGAAGCATGCGCATGATTCTTGAGATTATGCAAAAGAAGCTGATGAGACCA   820

Qy   878 AACTAGCCGAAGAGATTCCTCTGAAAATCTTGGCACCACAATGGCTTGGTTGGAAGACTGA   937
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   821 AACTGGCTGAGGAGGTTCCTCTGAAAATCCTGGCCCACAATGGCTTCGTTGGAAGACTGA   880

Qy   938 TTGGAAAAGAGGCCAGAAATTGAAGAAATTGAACATGAAACATGAAACAGGGACCAAGATAACAA   997
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   881 TTGGCAAAGAGGCCAGAAACCTGAAGAAATTAGAACATGAGAAAATAGAGGACCAAGATAACCA   940

Qy   998 TCTCATCTTTGCAGATTGAGCATTTATGGAAGCTGTGAAGGGCA   1057
          ||||||||||||||||||||||||||||||||||||||||||||
Db   941 TCTCATCCTTGCAGATTGAGCATTTATAACCCCGAGAGAACCATCACCGTGAGGGGCA   1000

Qy   1058 CAGTTGAGGCCTGCGTGTGCCAGTGCTGGAGATGAGAGAAGAAGCTGCGTGAGGCCTTTG   1117
          |||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db   1001 CCATTGAGCCTGCGCTGTGCCAATGCTGTGAGATAGAGATTATGAAGAAGCTCCCAGAGGGCCTTTG   1060

us-09-270-437d-6.rst

```
Qy   1118 AAAATGATATGCTGGCTGTTAACCAACAAGCCAATCTGATCCCAGAGGTTGAACCTTCAGCCG 1177
          |||||  || |||||||  ||||||||||||||||||||||||||
Db   1061 AGAACGACATGCTGGCCGTTAACCAACAAGCCAATCTGATCCGTGCTGTCTGCTTCTTCAGTG 1120

Qy   1178 CACTTGGCATCTTTCAACAGGACTGTCCGTGTATCTGCTTCCTCCA           1237
          |||||||||||||||    ||||  ||||||  |||||||||||
Db   1121 CACTTGGCATCTTTCATCTGTGTGCTGTCGTTCCTCCA-CAGCAGGCCCCGTGTGAG     1179

Qy   1238 CTCCCCCGGCTGCCCCCTACCACCCCACTCCGATACTTCTCCAGCCTGT          1297
Db   1180 TTCCCCCAGTCCTCCCTATCACCCCTTGCTACCCACTCCGATACTTCTCCAGTCTGT      1239

Qy   1298 ACCCCATCCACCAGTTGGCCCGTTCCGCCATCACTCTTATCCAGACCAGGAGGATTG      1357
Db   1240 ACCCTCACTCACCATTTCCGGCCCCATTCCCACCATCATCACTCTTACCCACCAGGAGGACTG 1299

Qy   1358 TGAATCTCTTCATCCCAACCCAGGCTGTGGGCGCCATCATCGGGAGGAAGGGGCACACA    1417
Db   1300 TAAGTCTCTTCATCCCAACCCAGGCTGTGTATCATCGGGAGGAAGGGGCACACA       1359

Qy   1418 TCAAACAGCTGGCCAGATTGCCCGGAGCCTCTATCAAGATTGCCCTGCGGAGGCCCAG     1477
Db   1360 TCAAACAGCTGGCTCGATTGCTGGTGGCTCCATCAAGATTGCTCCAGCAGAAGGTCCAG     1419

Qy   1478 ACGTCAGCGAAAGGATGTCATCATCACCGGGCCACCGAAGCCCAGTTCAAGGCCCCAGG    1537
Db   1420 ATGTCAGTCAGTGAGGATGGTCATCATCACTGGTCCTCCTGAAGCCCAGTTTAAGGCTCAGG   1479

Qy   1538 GACGGATCTTTGGGAAACTGAAAGAGGAGAAACTTCTTTAACCCCAAGAAGAAGTGAAGC   1597
Db   1480 GACGGATCTTTGGGAAACTGAAAGAGGAGAAACTTCTTTAATCCCAAGAAGAAGTGAAGC   1539

Qy   1598 TGGAAGCGCATATCAGAGATGCCCCTTCCACAGCTGGCGGGTGATTGGCAAAGGTGGCA    1657
Db   1540 TGGAAGCCGCCACATCCGAGTTCGTGTCGAACCCTGGCGGGTGATTGGCAAGGGCGGGA    1599

Qy   1658 AGACCGTGAACGAACTGCAGAACTGCAGAACTGCAGAAGTGAGAAGCTCATCGTGCCTCGTGACCAAA   1717
Db   1600 AAACCGTGAACGAGCGCTGCAGAACTGCAGAAGTGCAGAAGTGATCTGCTCGTGACCAAA    1659

Qy   1718 CGCCAGATGAAAATGAGGAAGTGATCGTCAGAATTCGGGCACTTCTTGCTAGCCAGA    1777
Db   1660 CGCCAGAACGAGAATGAGGAAGTGATCGTCAGAATTCGGGCCATTTTTTTGCTAGCCAGA    1719

Qy   1778 CTGCACAGCGCAAGATCAGGGAAATTGTACAACAGTGAAGCAGCAGCAGAAATACC     1837
Db   1720 CTGCACAAATCATACCGAGTTCGTACAACAGTGAAGCAGCAGCAGAGATACC       1779

Qy   1838 CTCAGGGAGTCGCCTCACAGCGCAGCAAGCGCAGCTCCACAGGCACCAGCAAAACAACG   1897
Db   1780 CTCAGGGAGTCGCCTCACAGCGCAGCAAGCGCAGCTCCCACAG--CACCAGCAAGCAACC    1837

Qy   1898 GATGAATGTAGCCCTTCCAACACTTGCACAGTGCGGAAGCAGCAGAAGATGAG       1957
Db   1838 GATGAATGTAGCCCTGCCTTCCAACACTGCACAGTGCGGAAGCAGCAGA-ATGAGAC     1893

Qy   1958 GAGCAAACCAAAGACCAT-CTGAGGAATGAGAAGTGCTGCGAGGGCGGCCCAGGGTTTGCCAG  2016
Db   1894 GAGCAAACCAAAGAGCAGCATCCCCCAGCTTCCCCCTGGCTTCTGCGAGGAGCAGGCAGATCCG  1950

Qy   2017 CGAGGCCCTGAGAACCCCAGGGCCCAGGGCCGAGAGGGCGGGGAAGGTCAGCCAGG      2076
Db   1951 --AGACCTCTCACACATCCTGGGATCCGGGCACAGGGCACAGGGGCAAG--GCCCAGGTTGTCCCAG 2005

Qy   2077 AACCACCG---AGCCCCGCGCTCCCCCCCGGCGTT--CAGCCATCCAC         2132
Db   2006 AAAACACCGCTTGGGCTTCGCCCCCAGCTTCGCTCTGGCTTCTGCAGGCATAACAGCCATCCAC 2065

Qy   2133 TTCACCCATCCACTCGGATCCTCTCCGACTCCCGACGCTATCCCTTTTAGTTGAACTA     2192
Db   2066 TGCCCCATCCATCCAGAT-TCTCCCTCCAGTTCCCAGGACGCCGTATCCCTTTCGGTTGAACTA  2124
```

```
Qy   2193 ACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGCACACCC----TTTTTCTGTGTGGCAAA 2248
Db   2125 ACATAGGTGAACATGCTCAAAGCCAAGCAAAATTCCTAGCCGTTCTTTCTTGTGTGTGGAAAG 2184

Qy   2249 TCGTCTCTGTACGTGTGTACATATTAGAAGGGAA-GATGTTAAGATATGTGGCCTGT     2307
Db   2185 TTGTCTCTGTACATGTATGTACATATCAGAGGTGTTAAGATAGAAGAACTTTG      2244

Qy   2308 GGGTTACACAGGGTGCCTGCAGCCGTGTAATATTTAGAAATAATAATATCAAATAACTCA   2367
Db   2245 GGGTTACACAGGGTGCCTGCAGCGTGTAATATTTAGAAATAATAATATCAAATAACTCC   2304

Qy   2368 ACTAACTCCAATTTTAATCAATTATTAATTTTTTTCTTTTAAAGAAAGCAGCAGGCT     2427
Db   2305 ACTAACAACAATTTTTACTCAGTGATTAATTTTGTTTTCTTTTGAAGAAAGCAGGCT     2364

Qy   2428 TTTCTAGACTTAAAGAATAAAGTCTTTGG-GAGGTCTCACGGTGTAGAGGAGAGCTTTG   2486
Db   2365 TTTCTAGACTTCCAAGAAGAATAAAGTATTGGGTGAGGTCCACTGGTATAGAAGGAACTTTG 2424

Qy   2487 AGGC-CACCCGCACAAAATTCACCCAGAGGAAATCTCGTCGGAAGGACACTCACGGCAG    2545
Db   2425 AGCCACCCACACAACAAAAATTCACTTGGAGGGGAAATCTTGTC-AAGGACACTTAATGGCAG   2483

Qy   2546 TTCTGGATCACCTGTGATGTCAACAAGAGGATACCGTCCTTGAAGAGAAACTCTG      2605
Db   2484 TTGTGGGTCACCTGTGTCTGTCAACAGAAGGGATACCGTCCTCTGAGGAGGGGCCCTAGGC    2543

Qy   2606 TCACTCCTCATGCCTGTCTAGCTGACTCATACACCCATTTCTT-----------     2645
Db   2544 CCTGGTCACCTGTTCTCCCACACCCATTGTCTTTCACAGGTTTTTAAACTGATG       2603

Qy   2646 TTGCTTCACAGGTTTAAACTGTTTTTGTTTTTGTCATATCTGTATATAATTCTCTGTCTCTC   2705
Db   2604 ATTTTGTTTGTTTGTTTTGTTTTTGTTTCGATATCGTATATAGTTCTGTCTCTCTCC-     2662

Qy   2706 TGTTTATCTCCCCTCCCCTCCCCTCTGCTGCTCCCATTCTTTTGAATTTCC        2765
Db   2663 -----TTATCTCCCTCCTCCTAACCTTCCTCTGTCTGTCTTTACCCAAAAATTCTTTTGAGTCCC 2717

Qy   2766 TCATCCCCTCCATCTCAATCCCGTATCTCACGCACCCC---CCCCCCCCCAGGCAAAGCAGT 2822
Db   2718 TTCGTTTCTTTTCTGTCTCCCGTATCTATGCACCCTCTCCCCCTCCCCTCCCCAGGCAAAGCAGT 2777

Qy   2823 GCTCTGAGTATCACATCACACTGGTTTTGCATAACCAAAAGGAACAAAAGCGAAACACACCTCAACT 2882
Db   2778 GCTCTGAATATCATCACACCACACCACCACACCGAACAACAAATGCAACCACAGACCAGCCTCAGCT 2837

Qy   2883 TACACTTGGTTACTCAAAAGAACAAGAGTCAATGTACGGGGTTTTGGAAGAG       2942
Db   2838 TACACTCCGTTACTCAAGCGAACCAAGAGTCGGCGGGGTACTTGTCCTCCTAGC-ATTCGGAAGAG 2896

Qy   2943 GAAAACAGGAACCCACCAAATCAACCAAAGAAAAATTCACACAATGAA        3002
Db   2897 GAAAGCAGGATCCCAGCAGAATAAACCAGCCAAACAAAG-AAAAATCCCCACAAAGAA      2955

Qy   3003 AGAATGTATTTTGTCTTTTGTTGTTTTGGTGTATAAGCCATCAATATTCAGCAAAAATGAT   3062
Db   2956 AGAATGGA----CTCTCTCCATCTTGGTATACAAGCCATCAACATTTAGTGAAATTCT      3011

Qy   3063 TCCTTTCTTTAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCCCA   3122
Db   3012 TTCTTTCTTTAAAAAATTTAAAGTGCAGGAAAATATGCAATTTATACGAGGTTGTTGGCCCA   3071

Qy   3123 GGGCGGTAAATTCACAGATTTTTTTAACAGAAAAAACACACGA---AGAAGTACCTCA    3179
Db   3072 GGGCGGTAAATTTACAGATTTTTTTTAAATGAGAAAAACACACAAATAAAAGTCACCTCA    3131

Qy   3180 GGGTG-TTTTTTACCTTGCTCAGCACCTTTTGTGTCTTTGTTTCCCTTAGAGATTTTGTAAAGCTGAT 3238
Db   3132 GGTGTTTTTTTACCTTGCTCTTCAGCACCGTATCCCTTTCGTTTCCCTTAGAGATTTTGTAAAACTGAT 3191

Qy   3239 AGTTGGAGCATTTTTTTATTTTTTTAATAAAAATGAGTTGGAAAAAATAAGATATCAA    3298
```

```
Db    3192  |||||||||  |||||||||
            AGTTCGGAGCA-TTTTTTATTTTTTAATAAAAATGAGTT-GAAAAAAAAAAAGATGAG 3249

Qy    3299  CTGCCAGCCTGGAGAAGGTGCAGTGCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCG 3358
            |||||||||||||||||  |||||  ||||||||||| |||||||      |||||||||
Db    3250  CTGCCAGCCCCAAGCAGCGGGCAGCCCAGGTGTGCCAGCAGCTCTTC-----TGTCTTCCG 3304

Qy    3359  CTAGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3412
            ||||||||||||| || |||||||||| |||||||||||||||  ||| |||||
Db    3305  CTAGCCAAGAGCCTACGTGCCTTCTTGTGGACAAACCTAAAAATGTTTATTTT 3358
```

RESULT 2
AK077404
LOCUS　　　　AK077404　　　　　　3262 bp　　　mRNA　　　　linear　　HTC 18-SEP-2003
DEFINITION　Mus musculus 6 days neonate head cDNA, RIKEN full-length enriched
　　　　　　　library, clone:5430439D08 product:HEPATOCELLULAR CARCINOMA
　　　　　　　AUTOANTIGEN homolog [Homo sapiens], full insert sequence.
ACCESSION　　AK077404
VERSION　　　AK077404.1　GI:26097330
KEYWORDS　　HTC; CAP trapper.
SOURCE　　　Mus musculus (house mouse)
　ORGANISM　Mus musculus
　　　　　　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
　　　　　　　Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE　1
　AUTHORS　Carninci,P. and Hayashizaki,Y.
　TITLE　　High-efficiency full-length cDNA cloning
　JOURNAL　Meth. Enzymol. 303, 19-44 (1999)
　MEDLINE　99279253
　PUBMED　10349636
REFERENCE　2
　AUTHORS　Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
　　　　　　Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
　TITLE　　Normalization and subtraction of cap-trapper-selected cDNAs to
　　　　　　prepare full-length cDNA libraries for rapid discovery of new genes
　JOURNAL　Genome Res. 10 (10), 1617-1630 (2000)
　MEDLINE　20499374
　PUBMED　11042159
REFERENCE　3
　AUTHORS　Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
　　　　　　Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
　　　　　　Sumi.N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
　　　　　　Yamamoto,R., Matsumoto,K., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
　　　　　　Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
　　　　　　Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
　　　　　　Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
　TITLE　　RIKEN integrated sequence sequence analysis (RISA) system--384-format
　　　　　　sequencing pipeline with 384 multicapillary sequencer
　JOURNAL　Genome Res. 10 (11), 1757-1771 (2000)
　MEDLINE　20530913
　PUBMED　11076861
REFERENCE　4
　AUTHORS　The RIKEN Genome Exploration Research Group Phase II Team and the
　　　　　　FANTOM Consortium.
　TITLE　　Functional annotation of a full-length mouse cDNA collection
　JOURNAL　Nature 409, 685-690 (2001)
REFERENCE　5
　AUTHORS　The FANTOM Consortium and the RIKEN Genome Exploration Research
　　　　　　Group Phase I & II Team.
　TITLE　　Analysis of the mouse transcriptome based on functional annotation
　　　　　　of 60,770 full-length cDNAs
　JOURNAL　Nature 420, 563-573 (2002)
REFERENCE　6　(bases 1 to 3262)
　AUTHORS　Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
　　　　　　Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
　　　　　　Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
　　　　　　Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
　　　　　　Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
　　　　　　Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
　　　　　　Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
　　　　　　Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
　　　　　　Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
```

```
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..3262
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:5430439D08"
                     /db_xref="MGI:2393565"
                     /db_xref="taxon:10090"
                     /clone="5430439D08"
                     /tissue_type="head"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="6 days neonate"
     misc_feature    16..1796
                     /note="HEPATOCELLULAR CARCINOMA AUTOANTIGEN homolog [Homo
                     sapiens] (SPTR|Q9Y6M1, evidence: FASTY, 88.3%ID,
                     100%length, match=1775)
                     putative"
     polyA_signal    3249..3254
                     /note="putative"
     polyA_site      3262
                     /note="putative"
ORIGIN
Query Match             58.8%;  Score 2006.2;  DB 11;  Length 3262;
Best Local Similarity   81.5%;  Pred. No. 0;
Matches 2693;  Conservative   0;  Mismatches  483;  Indels  127;  Gaps   27;

Qy    58  GGGAAGAGACGGATGATGATGAACAAGCTTACATCGGGAACCTGAGCCCCGGCCCCGTCACCGCC 117
          ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||
Db     4  GGGAAGAGACGGATGATGATGAACAAGCTGTCATTGGGAACCTGAGCCCCGGCCCCGTCACCGCC 63

Qy   118  GACGACCTCCGGCAGTCTTTGGGACACAGGAAGCTGCCCCTGGCCGGGACAGGTCCTGCTG 177
          ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    64  GACGACCTCCGGCAGTCTTCGGGGACACAGGAAGCTGCCCCTGGCCGGGACAGGTCCTACTC 123

Qy   178  AAGTCCGGCTACGCCTTCGGTGGACTACCCGACCAGAAGTGGGCCATCCGCGCCATCGAG 237
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   124  AAGTCCGGCTACGCCTTCGGTGGACTACCCGACCAGAAGTGGGCCATCCGCGCCATCGAG 183

Qy   238  ACCCTCCGGGTAAAGTGGAATTGCATGGAAGAGAAAATCATGGAAGTTGATTACTCAGTCTCT 297
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   184  ACCCTCCGGGTAAAGTGGAATTGCATGGAAGAAATCATGGAAGTTGACTACTCAGTCTCT 243

Qy   298  AAAAAGCTAAGGAGCAGGAGAAAATTCAGATTCGAAACATCCCTCCTCCACCTGCAGTGGGAG 357
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   244  AAAAAGCTAAGGAGCAGGAGAAGAATCCAGATTCGGAACATCCGCCTCACCTGCAGTGGGGAG 303

Qy   358  GTGTTGGATGGACTT-TTGGCTCAATATGGACAGTGGGAATGTGGAACAAGTCAACAC 416
          |||||||||||||||  |||||||||||||||||||||||||||||||||||||||||
Db   304  GTGTTGGATGGGCTTGTTGGCTGAATATGGGACAGTGGGAGCGTGGGAGCAAGTCAACAC 363

Qy   417  AGACACAGAAACCGCGTTGTCAACGTCACATATGCACAAGAAGAAGCAAAAATAGC 476
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   364  AGATACAGAAACTGCCGTTGTCAACGTCACCTATATGACCAAGAAGAAGAGGCAAGCCAGC 423
```

```
Qy   1556 TGAAAGAGGAAAACTTCTTTACCCCAAAGAGAAGTGAAGCTGGAAGCGCATATCAGAG 1615
          |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db   1483 TGAAGGAAGAAAACTTCTTTAATCCCAAAGAGAAGTGAAGCTGGAGGCCCACATCCGAG 1542

Qy   1616 TGCCCCTTCACAGTCTGGCGCGGGTGATTGGCAAAGGTGGCAAAGACCGTGAACGACTGC 1675
          |||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db   1543 TCCCCTGTCGACGCGCTGGCGGCCGGGTGATTGGCAAGGGCGGGAAAACCGTGAACGAGCTGC 1602

Qy   1676 AGAACTTAACCAGTGCGAGAGTCATCGTGCCTCGTCGCCAAACGCCAGATGAAAATGAGG 1735
          ||||||| ||||||||||||||  |||||||||||||||||||||||||||||||||||
Db   1603 AGAACTTGACAGTGCGAGAAGTTATCGTGCCTCGTCGACCAAACGCCAGAGAATGAGG 1662

Qy   1736 AAGTGATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCCAAGATCA 1795
          ||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||
Db   1663 AAGTGATCGTCAGAATTATCGGGCATTTTTGCTAGCCAGACTGCACAAGCGGCAAGATCA 1722

Qy   1796 GGGAAATTGTACAACAGTGAAGCAGCAGGAGCAGAGAAATACCCTCAGGGAGTCGCCTCAC 1855
          ||||||||||||||||  |||||||||||||||||||||||||||||||||||||||||
Db   1723 GGGAAATTGTACGCAGTGAAGCAGCAGGAGCAGAGATACCCTCAGGGAGTCGCCCCAC 1782

Qy   1856 AGGCGCAGCAAGTGAGGCTCCCACAGCGCACCCAGGACTTCTGAAGCATGCCG 1915
          ||||||||||||||||||||||   |||||||||||||||||||||||||||||
Db   1783 AGGCGCAGCAAGTGAGGCTCCCACAG--CACCAGCAAGCAAGCCAGATCGGGAGCAAACCAAAGACCAT 1840

Qy   1916 AACACCTTGACAGAATGACACAAACGCAGCCCAGATCGGGAGCAAACCAAAGACCAT 1975
          |||||||||||||| |    |||||||||
Db   1841 AACACCTGACAG-ATGAGACCAAAC--AGCCAGATCGGGAGCAAACCAAAGAGCAT 1896

Qy   1976 -CTGAGGAATGAGAAGTCTGCGGCGGCCGCCCAGGGACTCTGCGAGGACCACACG--AGCCCCGC 2034
           | |||||||||||||||    ||||||||||||||||||||  |
Db   1897 CCCGAGGAGTGCGCAGTCTGCAGAGCACAGGGGCCAGGGGCCT----GCAGACCTCTACACATCC 1952

Qy   2035 AGGGGCCGGAGGAGGGGCGGGGAAGTCAGCCAGTTTGCCAGGATTT--CAGCCATCCACTTCACCATCCCCATCTCGGAT 2092
          |||||||||||||||||||    ||||||||||||||||  |
Db   1953 TGGGATTCAGGAGGGCACACAGGGAAG---GCCAGTTGTGTCCAGAAACACCGCGCTTGGCCTGC 2009

Qy   2093 CTCCCGCCCCCCAGGGCTTCTGCGAGGCTT--CAGCCATCCACTTCACCATCCCCATCTCGGAT 2150
          | ||||||||||||||||||||||||||   ||||||||||||||||  |||||
Db   2010 CCCCAGCTTCCCCTGGCTTCTGCGAGGCATAACAGCCATCCCATCGCCATCTCAGAT 2069

Qy   2151 CTCTCCCTGAACTCCCACGACGCTATCCCTTTAGTTGAACTAACATAGGTGAA--CGTGT 2208
          |||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db   2070 -TCTCCTCAGTTCCCAGACGCTATCCCCTTTCGTTGAACTAACATAGTGAAACATGCT 2128

Qy   2209 TCAAAGCCAAGCAAAATGCACACCCTTTTC-----TGTGGCAAATCGTCTCTGTACATG 2263
          ||||||||||||||||||||||||||||||
Db   2129 CAAAGGCCAAGAAAAATCCTAGCCGTTCTTTCTTTGTGGAAAGTTGTCTCTGTACATG 2188

Qy   2264 TGTGTACATATTAGAAAGGGAA-GATGTTAAGATAGATAGTGGCCTGTGGGTTACACAGGGTG 2322
          |||||||||||||||||||||  ||||||||||||||||||  |||||||||||||||||||
Db   2189 TATGTACATATCAGAAGGGGATGTTAAGAAATGTTAAGAAATGTGGCCTGTGGGTTACACAGGGTG 2248

Qy   2323 CCTGCAGCGCGTAAATATATTTAGAAATATATCAAATAATCAAATTCTTTAAA 2382
          |||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db   2249 CCTGCAGCGTTAATATCAGAAGGGGGATGGGAAGAGCAGCAGAGGCTTTTCTAGACTTCCAA 2308

Qy   2383 TAATCAATTATTAATTTTTTTTCTTTTTGAAGGAGGTCTCACGGT------GTAGAGGAGCTTTGAGGCCACC 2442
          ||||||||||||||||||||  ||||||||||||||||||
Db   2309 TACTCAGTGATTAATTTTG-TTTCTTTTTGAAGAGAAAGCAGGCTTTTCTAGACTTCCAA 2368

Qy   2443 GAATAAAGTCTTTGGAGGTCTCACGGT------TTCACCCAGAGGAAATCTCGTCGGAA 2494
          |||||||||||||||||||||||||||
Db   2369 GAATAAAGATTGGGTGAGGGTCCCACTTGGTATTAGAAAAGGAACTTTGAAGCCCACC 2428

Qy   2495 CGCACAAAA--------TTCACCCAGAGGAAATCTCGTCGGAA 2530
          ||||||||||
Db   2429 CACACAAAAAAAATTTCACTTGGAGGACCACCTTTAATTGCCC 2488

Qy   2531 GGACACTCACGGCAGTTCTGATCACCTGTATGTCAACAGAAGGGATACCGTCTCCTT 2590
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2489 AGTTTGTTGGGTTCCACCTGTTTCGTCCAAGCCAGAAAAGGGAAAAACCCTTCCCT 2548

Qy   2591 GAAGAGGAAACTCTGTCACTCCTCACTGTCTCATGCCTCTGTAGCTCATACACCCATTTCTTGCT 2650
```

```
Qy   477 CATGGAGAAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCC 536
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   424 TATTGAGAAGCTCAGTGGGCATCGGTCATCAGTTTGAGGACTACTCCTTCAAGATTCCTACATCCC 483

Qy   537 GGATGAAGAGGTGAGCTCCCCTTCGCCCCCTCAGCGCGAGCCCACTCTTC---------TGC 596
         ||||||||||||||||||||  |||||||||||||||||||||||||
Db   484 CGATGAAGAGGTGAGCTCTCCTTCACCCCCTCATCG--------------TGC 522

Qy   597 CCGGGACCAAGGCCACAGCGCCCGTGGGGCACTTCTCAGGCCAGACGATTGATTTCCCGCT 656
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   523 CCGGGAACCAAGGCCACAGCGGCCCGGGGAGTCTTCTCAGGCCAGACAGATTGATTTCCCGCT 582

Qy   657 GCGGATCCTGGTCCCCACCCAGTTTGTTGGTGCCCATCATCGGAAAGGAGGGGCTTGACCAT 716
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   583 GCGGATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGGCTTGACCAT 642

Qy   717 AAAGAACATCACTAAGCACAACCAGTCCCGGGTAGATATCCATCATAGAAAAGAGAGAACTCTGG 776
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   643 AAAGAACATCACTAAGCACAGACCCGGGTAGACATTCCACAGAAAGGAGAGAACTCTGG 702

Qy   777 AGCTGCAGAGAAGCCGTGTACCATCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCG 836
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   703 GGCTGCGAGAAGCCGTGTACCATCATGCCTACCCCAGAAGGGGACATCTGAAGCATGCCG 762

Qy   837 CATGATTCTTGAAATCATGCAGAAAGAGCAGATGAGACCAAAAACTAGCCGAAGAGATTCC 896
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db   763 CATGATTCTTGAAATCATGCAAAATGGCTTCGTTGAGACCAAACTGGCTGAGGAGGTTCC 822

Qy   897 TCTGAAAATCTTGGCACACAATGGCTTGGTTGGTGGAAGACTGATTGTGGAAAGAAGGCAGAAA 956
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   823 TCTGAAAATCCTGGCCCAAATGGCTTCGTTGGAAGACTGATTGGCAAAGAAGGCAGAAA 882

Qy   957 TTTGAAGAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTT 1016
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   883 CCTGAAGAAAATAGAACACATAGAGAGAGAATCTCGCAGAGATAACCAGAGGCAGAAA 942

Qy   1017 GAGCATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACACTGTGAGGCCTGTG--CCA 1075
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   943 GAGCATTTATAACCCGAGAAGAACCATCACCGTGAGGGGAGGGGCACCATTGAAGCCTGTGCCCA 1002

Qy   1076 GTGCTGAGATAGAGATTATGAAGAGAGTCTGCTGAGCCTTTGAAAATGATATGCTGGCTG 1135
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1003 ATGCTGAGATAGAGATTATGAAGAGAGTCTCCGAGAGGCCTTGAGAACGACATGCTGGCCG 1062

Qy   1136 TTAACCAACAGAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCGCACTTGGCATCTTTTCAA 1195
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1063 TTAACCAACAGCCAATCTGATCCCAGGGCTAAACCTCAGTGCACTTGGCATCTTTTCGA 1122

Qy   1196 CAGGACTGTCCGTCGTATCTCTCCACCACTCCGGATACTTCTTCAGAGCAGGAGTTGTCTCCCCGTCCCCCT 1255
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1123 CTGGACTGTCTGTGTCTCCTCCACCACTCCGGATACTTCTTCCTCACAGTGAGTTCCCCCAGTCTCCCCT 1182

Qy   1256 ACCACCCCTTCACTACCACCCATCATCACTCTTATCCGAGCAGCAGATTGTGAATCTCTTCATCCCAA 1315
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   1183 ATCACCCCTTGCTACCACCCACCATCATCACTCTTCTCCGAGCAGACTGTGAATCTCATCATCCCAA 1242

Qy   1316 GCCCGGTTCCCGCATCATCAGTCCGGAAGGCCACACATCAAACAGCTGCGAAAGCAGTCGCTCGGAT 1375
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1243 GCCCATTCCCAGTCTGGTGGGGCCATCATCGGGAAGAGGGGGGCCACACATCAAACAGCTGTAAGCTCGGAT 1302

Qy   1376 CCCAGGGCTGTGGGAGCCTCTATCCGAGATTGCCCCTGCGGGAAGGCCCCAGACGTCAGCGGAAGGATGG 1435
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1303 CCCAGGGCTGTGGGATCATCAGTTGCCTGCGGGAAGGCCCCAGACGTCAGTGAGAGGATGG 1362

Qy   1436 TCGCCCGGAGCCTCTATCAAGATTGCCCCATCCGGGACACCTCAGGCCCCAGGCGATCTTTGGGAAAC 1495
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1363 TTGCTGGTGCTCATCCAATCAAGATTGCTCTCCAGCAGACTCCAGATGTCAGTGAGGAGGATG 1422

Qy   1496 TCATCATCACCCGGGCCACCGGGCCACTCTGGGGAAGCCCAGTTCAAGGCCCAGGACGATCTTTGGGAAAC 1555
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1423 TCATCATCACTGTCCGTCCTCCTGAAGCCCAGTTTAAGGCTCCGAAGCCGGATCTTTCGGAAAC 1482
```

```
Db  2549 TAAGGGAAGGGGCCCTAGGCCCTTGTCACCCGTTTCTCCCACACCCATTGCTCAGTT 2608
              ||| ||| |   || | |                   |||| ||||  || |
Qy  2651 TCACAGGTTTTAAAC-----------------------TGGTTTTTTGCATACTGCT 2684
         |||||||||                               |||||||||||||||||
Db  2609 TCACAGGTTTTTAAATTAAATAATTTTGGTTTGTTTTTGCATACTGCT 2668

Qy  2685 ATATAATTCTCTGTCTCTCGTTTATCTCCCCCCCCCCCCTTCTTCTCCA 2744
         |||||                | |||||||| | |||| |||
Db  2669 ATATAGT--TCTGTCTCTCTCTCTTATCTCCCTCCTAACCTTCCTCTGTCTTCTTTACCC 2726

Qy  2745 TCTCCATTCTTTTGAATTCC---TCATCCTCCATCCAATCCCGTATCACCGACACC-- 2799
         |||                   ||| |  | |  | | |   | | |||| |
Db  2727 CAGAAATTCTTTTTGAGTCCCCTTTTCGTTCTTTTTTCGCTCCCCGTATCTATGCACCCC 2786

Qy  2800 -CCCCCCCCCCAGGCAAAGCAGTGCTCTGAGTATCACATCACAAAAGGAACAAAAGC 2858
          ||||                 ||| | | | |  || ||| |   || | |||||
Db  2787 TCCCCTCCCCCAGGCAAAGCAGTGCTCTGAATATCATACCACACAAGGAACAAATGC 2846

Qy  2859 GAAACACACAAACCAGCCTCAACTTACACTTGGTTACTCAAAAGAACAGAGTCAATGGT 2918
         |                ||| |||| |||||||| |||  ||| || |  | |  |
Db  2847 GAACCACCAGACCAGCCCAGCTTACACTCGGTTACTCACAAGCGAACAGAGTCGGCGGT 2906

Qy  2919 ACTTGTCCTAGCCGTTTTGGAAGAGAGAAAACAGAAACCCACCAAATCAACCAA 2978
         ||||| |||||         ||| |||||||| | |   |||| | ||||| |
Db  2907 ACTTGTCCTAGC-ATTCGGAAGAGAGAGAAGCAGGATCCCAGCAAATAAACCAGCCAGCCAA 2965

Qy  2979 ACAAAGAAAAAATTCCACAATGAAAGAATGTATTTGTCTTTTGCATTTTGGTGTATAA 3038
         ||||| |||||           ||||| |   || |  |  ||||||||| | ||
Db  2966 ACAAAG-AAAAATCCCACAAAGAAAGAATGGA----CTCTCTCCCATCTTGGTATACAA 3020

Qy  3039 GCCATCAATATTCAGCAAAATGATTCCTTCTTTAAAAAAATGTGGAGGAAAGTAG 3098
         ||||| ||||||| | |  |  |  ||| |  ||| |
Db  3021 GCCATCAACACTTTAGTGAATTCTTCTTCTTTAAAAATTTAAAGTGCAGGAAAATAGC 3080

Qy  3099 AAATTTACCAAGGTTGTTGGCCCAGGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAA 3158
         ||| |||||||||| ||||| | ||||| ||||| | | |||| |||||  |||||
Db  3081 AAATTTATACGAGGTTGTTGGCCCAGGGCCGTTAAATTTACAGATTTTTTTAAATGAGAAA 3140

Qy  3159 CACACAGA---AGAAGCTACCTCAGGTG-TTTTACCCAGCACCTTGCTCTTGTGTTTC 3214
         |||| ||    ||| ||||||| |  |  ||||| |||||| || |||| ||| |||
Db  3141 AACACACAAATAAAAGCTACCTCCCAGGTGTTTTTTTACCCTCAGGCACCTTGCTCTTGTGTTTC 3200

Qy  3215 CCTTAGAGATTTTGTAAAGCTGATAGTTGGAGCATTTTTTTTATTTTTTTAATAAAAATGA 3274
         |||||||||||| || |   |  | |   ||||||||   | || || ||||| |||||
Db  3201 CCTTAGAGATTTTGTAAAACTGATAGTTGGAGCA-TTTTTTATTTTTTTTAATAAAAATGA 3259

Qy  3275 GTT 3277
         |||
Db  3260 GTT 3262
```

RESULT 3
AK044984
LOCUS   AK044984   2964 bp   mRNA   linear   HTC 20-SEP-2003
DEFINITION Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
            enriched library, clone:B130017M24 product:HEPATOCELLULAR CARCINOMA
            AUTOANTIGEN homolog [Homo sapiens], full insert sequence.
ACCESSION   AK044984
VERSION   AK044984.1 GI:26090647
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,

        TITLE   Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
                    Normalization and subtraction of cap-trapper-selected cDNAs to
                    prepare full-length cDNA libraries for rapid discovery of new genes
        JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
        MEDLINE   20493374
         PUBMED   11042159
    REFERENCE   3
        AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
                    Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
                    Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
                    Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
                    Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
                    Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
                    Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
        TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
                    sequencing pipeline with 384 multicapillary sequencer
        JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
        MEDLINE   20530913
         PUBMED   11076861
    REFERENCE   4
        AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
                    FANTOM Consortium.
        TITLE   Functional annotation of a full-length mouse cDNA collection
        JOURNAL   Nature 409, 685-690 (2001)
    REFERENCE   5
        AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
                    Group Phase I & II Team.
        TITLE   Analysis of the mouse transcriptome based on functional annotation
                    of 60,770 full-length cDNAs
        JOURNAL   Nature 420, 563-573 (2002)
    REFERENCE   6 (bases 1 to 2964)
        AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
                    Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
                    Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
                    Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,T., Kasukawa,T.,
                    Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
                    Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,T., Murata,M.,
                    Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
                    Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
                    Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
                    Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
                    Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
                    Muramatsu,M. and Hayashizaki,Y.
        TITLE   Direct Submission
        JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
                    Physical and Chemical Research (RIKEN), Laboratory for Genome
                    Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
                    RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
                    Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
                    URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
                    Fax:81-45-503-9216)
        COMMENT   cDNA library was prepared and sequenced in Mouse Genome
                    Encyclopedia Project of Genome Exploration Research Group in Riken
                    Genomic Sciences Center and Genome Science Laboratory in RIKEN.
                    Division of Experimental Animal Research in Riken contributed to
                    prepare mouse tissues.
                    Tissues were provided by Dr. Tomohiro Kono (Department of Animal
                    Science, Tokyo University of Agriculture, 1737 Hunako Atsugi City,
                    Kanagawa Prefecture, Japan) whose assistance we gratefully
                    acknowledge.
                    Please visit our web site for further details.
                    URL:http://genome.gsc.riken.go.jp/
                    URL:http://fantom.gsc.riken.go.jp/.
        FEATURES   Location/Qualifiers
        source   1..2964
                        /organism="Mus musculus"
                        /mol_type="mRNA"
                        /strain="C57BL/6J"
                        /db_xref="FANTOM_DB:B130017M24"
                        /db_xref="MGI:2410341"
                        /db_xref="taxon:10090"
                        /clone="B130017M24"
                        /tissue_type="parthenogenote"

misc_feature

/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="9.5 days embryo"
1..2964
/note="HEPATOCELLULAR CARCINOMA AUTOANTIGEN homolog [Homo sapiens] (SPTR|Q9Y6M1, evidence: FASTY, 88.3%ID, 100%length, match=1775)"

ORIGIN

Query Match   54.6%;   Score 1862.8;   DB 11;   Length 2964;
Best Local Similarity 81.9%;   Pred. No. 0;
Matches 2441;   Conservative 0;   Mismatches 452;   Indels 87;   Gaps 22;

```
Qy    247 GGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTA 306
          ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     30 GGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGACTACTCAGTCTCTAAAAAGCTA 89

Qy    307 AGGAGCAGGAAAATTCAGATTCGAAACATCCCTCCACCTGCAGTGGGAGTGTTGGAT 366
          ||||| || |||||||||||||||||||||||||||||||||||||||||||||||
Db     90 AGGAGCAGAAGAATCCAGATTCGGAACATCCCTCCACCTGCAGTGGGAGTGTTCGAT 149

Qy    367 GGACTTTTGGCTCAATATGGACAGTGGAACAGTGAACAAGTCAACACAGACACAGAA 426
          |||  ||||||||||||||||||||| ||||| |||||||||||||||||||||||
Db    150 GGGCTGTTGGCTGAATATGGGACAGTGGAGACGTGAGCAAGTCAACACAGATACAGAA 209

Qy    427 ACCGCCGTTGTCAACGTCACATATGCACATATGCCTATATGCAACAGAGAAGAAGCAAAAATAGCCATGGAGAAG 486
          ||||||||||||||||||||| ||||||||||||||||||||||||||||||||| ||||
Db    210 ACTGCCGTTGTCAACGTCACCTATGCAACTATGCCTATATGCAACAGAGAAGAAGCAGGCAAAGCTAGCTATTGAGAAG 269

Qy    487 CTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCTACATCCCGATGAGAG 546
          |||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db    270 CTCAGTGGGCATCAGTTTGAGGACTACTCCTTCAAGATTTCCTACATCCCGATGAGAG 329

Qy    547 GTGAGCTCCCCTTCGCCCCCCAGCGAGCCCAGCCGCCGAGCCCGGGGACCCTCCCGGGACCAA 606
          |||||||||||||| |||||||||||||||||||         ||||||||||||||
Db    330 GTGAGCTCTCCTTCACCCCCCATCG------------------TGCCCGGGGACCAA 368

Qy    607 GGCCACGCCCTGGGGGCACTCTTCAGGCGACAGAGATTGATTTCCCGTGCCGATCCTG 666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    369 GGCCACGCCCGGGGGAGCCTCTTCAGGCGACAGAGATTGATTTCCCGTGCCGATCCTG 428

Qy    667 GTCCCCAGTTGTTGTTGGTGCCATCATCGGAAAGGAGGGGCTTGACCATAAAGAACATC 726
          ||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db    429 GTCCCCAGTTGTTGTTGGTGCCATCATCGGAAAGGAGGAGCTTGACCATAAAGAACATC 488

Qy    727 ACTAAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGAGCTGCAGAG 786
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    489 ACTAAGCAGACCCAGTCCCGGGTAGACATCCACAGAAAAGAGGAACTCTGGGGCTGCAGAG 548

Qy    787 AAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCCATGGCCGCATGATTCTT 846
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    549 AAGCCTGTCACAATCCATGCCACCCCAGAGGGGACTTCTGAAGCCATGGCCGCATGATTCTT 608

Qy    847 GAAATCATGCAGAAAGAGCAGATGAGAAGGCAGACCTAGCCGAAGAGATTCCTCTGAAAATC 906
          |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db    609 GAGATTATGCAAAAAGAGCGATGAGTGGAAGACAAACTAGCCGAGGAGGTTCCTCTGAAAATC 668

Qy    907 TTGGCACACAATGGCTTGGTTGGAAGCTGATTGGAAAGAAGGCAGAAATTTGAAGAA 966
          |||| ||||| |||||||||||||||||||||| ||||||||||||||||||||||
Db    669 CTGGCCCACAATGGCTTCGTTGGAAGCTGATTGCCAAAGAAGAAGGCAGAAACCTGAAGAA 728

Qy    967 ATTGAACATGAAAACAGGGACCAAGAAACATCTTTGCAGGATTTGAGCATATAC 1026
          |||||| ||||||||||||||||||||||| ||||||||||||||||||||||
Db    729 ATAGAACATGAGACAGGGGAAGCCAAATGATAACAATCCATCTGCCCATCGCAGGATTTGAGCATTTAT 788

Qy   1027 AACCCGGAAAGAACCATCACTGTGGAAGGCACAGTTGAAGGCACAGTTGAGGCTGTGCAGAGTGAGATA 1086
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    789 AACCCCGAGAGAACCATCACCGTGGAGGGCCTTTGAAAATGATATGCTGGCTTGTCAAAGCCAAGCA 848

Qy   1087 GAGATTATGAAGAAGTGCTGTAGGAGCTCCGAGAGGCCTTGAGAACGACATGCCCGATCCTTTTCAACAA 1146
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    849 GAGATTATGAAGAAGTGCCGAGAGGCCTTTGAGAACGACATGCCCGTTAACCAACAA 908

Qy   1147 GCCAATCTGATCCAGGGTTGAACCTGCACTTGGCATCTTTCACAGGACTGTCC 1206
          ||| ||||||||||||||||||||||||||||| |||||||
Db    909 GCCAATCTGATCCAGGGTCTAAACCTCCAGTGCACTGCACTCCCGCCCTACCACCCTTC 968

Qy   1207 GTGCTATCTCCACCAGCAGACTTCTCCAGCTGCTGTACCCGTACCCCATCACCAGTTGGCCGATCTTGG 1266
          |||| ||||||||| ||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    969 GTGCTTCCTCCTCCACCAGCAGCAGGGCCCCGTGGCAGTTCCCCGTACCCCAGTCCTCCCTATCACCCCTTT 1028

Qy   1267 ACTACCCACTCCGGATACTTCTTCCAGCGTGTGCACCCGCGATCACCAGTTGGCCCATCACC 1326
          |||||||||||||||||||||||||||| ||| ||||||||||||||||||||||| ||||
Db   1029 GCTACCCACTCCGGATACTTCTTCCAGTCTGTACCCTCATCACATTTCGGCCCATTCCAA 1088

Qy   1327 CATCATCACTCTTATCCAGACCAGGAGGATTGTGAATCTCTTCATCCAACCCAGGCTGTG 1386
          |||||||||||| |||| ||||||| |||||| |||||||||||||||||||||||||||
Db   1089 CATCATCACTCCTACCCAGACCAGGAGGACTGTAAGTCTCTTCGTCCCAACCCAGGCTGTG 1148

Qy   1387 GGCGCCATCATCGGGGGAACGGGGCCACACAAACAGTCGGCGAGATTGCCGGAGCC 1446
          ||||||||||||||| |||| |||||||||||||||||||||| ||||||||||||
Db   1149 GGTGCTTATCATCGGGGGAAGGGGCCACACAAACATCAAACTCGCTGATTTGCTGTGCC 1208

Qy   1447 TCTATCAAGATGCCCCTGCCGAAGGCCCAGACGTCAGCGAAAGGAGGATGGTCATCATCACC 1506
          ||| |||||||||||||||| |||||||||||| |||||||||||||||||||||||||||
Db   1209 TCCATCAAGATTGCTCAGCAGAAGGTCCAGATGTCAGTGTCAGTGGTCATCATCACT 1268

Qy   1507 GGGCCACCGGAGGCCCAGTTCAAGGGCCCAGGCCCCAGGCCGGACGATCTTGGGAAACTGAAAGAGGAA 1566
          |||| |||||||||||||||||||| ||||||| |||||||||||||||||||||||| |||||||
Db   1269 GGTCCCTCCGAGGCCCAGTTTAAGGGTCAGGGACCGGACATCTTTGGGAAACTGAAGGAAGAA 1328

Qy   1567 AACTTCTTAACCCCAAAGAAGAAGTTGGCAAAGGTGGCAAGACCGTGAACGAACTGCAACTTAACC 1626
          |||||| ||||||||||||||||||||||||||||||||| ||||||||||||||||||||| ||
Db   1329 AACTTCTTAATCCCAAAGAAGAAGTTGAAGGTGGCAAAGGGCCCACCATCCCGATGAACGAACTGCAACTTGACA 1388

Qy   1627 ACAGCTGCCCGGTGATTGGACAAGTGGCAAGACCGTGAACGAACTGCAGAACTTAACC 1686
          ||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   1389 ACCGCTGCCCGGTGATTGGCCTGACCAAAACGCCAGAGCGCAAGAGTGAAGTGATCGTC 1448

Qy   1687 AGTGCAGAGTCATCGTCGTCCTGTGACCCAGCGCACAGCAGCGCAAGATCAGGGAAATGTGATCGTC 1746
          ||||||||||||| ||||| |||||||||||||||||||| |||||||||||||| ||||||||||
Db   1449 AGTGCAGAAGTTATCGTGCCTGTGACACTCTGTACGTGCACAAGGCGCACTGAGAAGTGGAAATTGTA 1508

Qy   1747 AGAATTATCGGGACACTTCTTCTCAGGCGACAGAGATTGATTTCCCGTGCCGATCCTG 1806
          ||||||||| |||||||||||||||||||||||| ||||||| |||||||||||||||
Db   1509 AGAATTATCGGGCATTTTTTTCGTGCCTGTGACAGTGACTGACTGCAAGGACTGCAAGATCAGGGAATTGTA 1568

Qy   1807 CAACAGTGAAGCAACAGCAGCAGAAATACCCTCAGGGAGTCGCCTCACAGGGGCCAGCAAG 1866
          |||| ||||||||||||||||||||||||||||||||||| || ||||||||||||||||
Db   1569 CAGCAGTGAAGCAACAGCAGCAGAGATACCCTCAGGGAGTCGCCTCACAGGGGCCAGCAAG 1628

Qy   1867 TGAGGCTCCACAGGCACCAGCTTCTGAAGCATGCCGATGATTCTT 1926
          ||||||||||||| |||||||||| ||||||||||||||||||||
Db   1629 TGAGGCTCCCACAG--CACCAGCAAGCAAGCCATGATGATTGAATGATTCCCAACACCTGACA 1686

Qy   1927 GAATGAGACCAAACGCAGCCAGCCAGATCTGGAGCAAACCAAAGACCAT-CTGAGGAATG 1985
          ||||||||||||| |||||||||||||||||||||||||||||| |||| |||||||||
Db   1687 G-ATGAGACCAAAC---AGCCAGCAGATAGGGAGCAAAACCAAAGAGCATCCCCAGGAGTG 1742

Qy   1986 AGAAGTTCTGCGGAGCGGCGGCCGGAGGACTCTGCGCAGAGAACCCGAGGGGCCGAGG 2045
          || ||| ||||||||||| |||| |||||||||||||||| |||||||||||||||||
Db   1743 CGCAGTCTGCAGAGCAGCAGGCAGGCCGTGC-----AGACCTCTACACATCCTGGGATCCAGG 1797

Qy   2046 AGGGGCGGGGAAGGTCAGCCAGTTGCCAGAACCACCG--AGCCCCGGCTCCCGCCCCC 2103
          |||||||||||||| |||||||||| ||||||||||  ||||||||||| |||||||
Db   1798 AGGGGCACAGGGAAG---GCCAGTGTGTCCAGAAACACCGCTTGCCGCTGCCCCCAGCTTCC 1854

Qy   2104 CAGGGCTTCTGCAGGCTT--CAGCCATCCACTTCCACCATCCGGATCTCTCCTGAAC 2161
          |||||||||||||||    |||||||||||| |||||||||||||||| |||||||
Db   1855 CCTCGGCTTCTGCAGGCATAAACAGCCATCACTGCCCATCCAACTCAGAT-TCTCCTCAGT 1913

Qy   2162 TCCCCGACGCTATCCTTTTAGTGACTAACATCGGTTAACAGGCTGTTCAAAGCCAAGCA 2221
          |||||||||||||||||||||||||| |||||  |||||||| |||||||||||||||
Db   1914 TCCCGGCTTCTGCCATCCCTTTTCGGTTGAACTAACAAGCTCAAAGCAAAGCA 1973

Qy   2222 AAATGCACACCC-------TTTTCGTGGCAAATCGTCTCTGTGTACATGTTAG 2277
          |||| |||||||       |||
Db   1974 ...
```

```
VERSION      BQ691878.1  GI:21817194
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 874)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM2362  row: h  column: 03
             High quality sequence stop: 653.
FEATURES             Location/Qualifiers
     source          1..874
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6207026"
                     /tissue_type="ductal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC 110"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."

ORIGIN

Query Match             24.7%;  Score 842.2; DB 13;  Length 874;
Best Local Similarity   99.4%;  Pred. No. 2.3e-154;
Matches  856;  Conservative    0;  Mismatches   3;  Indels   2;  Gaps   1;
```

```
Qy   1824 GGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAGCAAGTGAGGCTCCACAGGCA 1883
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAGCAAGTGAGGCTCCACAGGCA 60

Qy   1884 CCAGCAAAACAACGGATGAATGTAGCCCTTCCACACCTGACAGATGAGACCAAACGCA 1943
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCAGCAAAACAACGGATGAATGTAGCCCTTCCACACCTGACAGATGAGACCAAACGCA 120

Qy   1944 GCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAGGAATGAGAAGCTCGCGGAGGCGG 2003
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAGGAATGAGAAGCTCGCGGAGGCGG 180

Qy   2004 CCAGGGACTCTGCCGAGGGCCCGAGACCCCAGGGGCCGAGGAGGGGGGAGGGGGGAGGTCAG 2063
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CCAGGGACTCTGCCGAGGGCCCGAGACCCCAGGGGCCGAGGAGGGGGGAGGGGGGAGGTCAG 240

Qy   2064 CCAGGTTTGCCAGAACCACCGAGCCCCGCCTCCGCCCCCCGCCGCCTCCGAGGCTTCG 2123
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CCAGGTTTGCCAGAACCACCGAGCCCCGCCTCCGCCCCCCGCCGCCTCCGAGGCTTCA 300

Qy   2124 GCCATCCACTTCACCATCCACTCCGGATCTCTCCTGAACTCCCGAGCCTATCCCTTTA 2183
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCCATCCACTTCACCATCCACTCCGGATCTCTCCTGAACTCCCGAGCCTATCCCTTTA 360

Qy   2184 GTTGAACTAACATAGGTGAACGTCTTCAAAGCCAAGCAAAATGCACACCTTTTCTGTG 2243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGCACACCTTTTCTGTG 420

Qy   2244 GCAAATCGTCTGTACATGTGTGTACATATTAGAAAGGGAAGATGTTAAGATAATGTGGC 2303
```

```
Db   1974 AAAATCCTAGCCGTTCTTGTTGTGGAAAGTGGTCTCTGTACATGTATGTACATATCAG 2033
          |||||||||||||||
Qy   2278 AAAGGGAA-GATGTTAAGATAATGTGGCCTGTGGGTACACAGGGGTGCCTGCAGCGGTAAT 2336
          ||| |||| ||||||||||||||||||||| |||||| |||||
Db   2034 AAGGGGAAGAAGGATGTTAAGATAATGTGGCCTGTGGGGTTACACAGGGGTGCCTGCAGCGGTTAAT 2093
Qy   2337 ATATTTTAGAAATAATATATATCAAATAACTCCAATTTTAATCAATTATTAA 2396
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   2094 ATATTTTAGAAATAATATATATCAAATAACTCCAATTTTACTCAGTGATTAA 2153
Qy   2397 TTTTTTTTCTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGATAAAAGTCTTTG 2456
          |||| |||||||||||||||||||||||||||||| |||||||||||||||||||
Db   2154 TTTTGTTTTCTTTTTGAAGAGAAAGCAGGCTTTTCTAGACTTCAAGAATAAAGTATTGG 2213
Qy   2457 G-GAGGTCTCACGGTGTAGAGAGAGCTTGAGGC-CACCCGCACAAAATTCACCCAGAG 2514
          |||||||||  |||      |||       |  |||||  ||
Db   2214 GTGAGGTCCACTGGTATAGAAAGGAACTTTGAGCCACCCACACAAAAATTCACTTGGAG 2273
Qy   2515 GGAAATCTCGTCGGAAGGACCAGTTCTGGATCACCTGTGTATGTCAACAGAA 2574
          ||||| |||  ||  |||     |||||||||||||||
Db   2274 GGAAATCTTGTC-AAGGACACTTAATGGCAGTTGTGGGGTCACCTGTCTGTCAACAGAA 2332
Qy   2575 GGGATACCGTCTCCCTGAAGAGGAAACTCTGTCACTCCCTCATGCCTGCTAGTCATACA 2634
          |||||||  | |     ||   ||       |
Db   2333 GGGATACCGTCCCCTGAGGGCCCCTAGGCCCTGGTCACCTGTTCTCCCACACACCCA 2392
Qy   2635 CCCATTTCTCTT------------TGCTTCACAGGTTTTAAACTGGTTTTTT 2674
          ||||||||||||            ||||||||||||||||||||||||||
Db   2393 TTGCTCAGTTCACAGGTTTTTAAACTGATGATTTTGGTTTGTTTTGTTTTTT 2452
Qy   2675 GCATACTGCTATATAATTCTCTGTCCTCTCTGTTATCTCTCCCCCTCCCCCTCCCC 2734
          |||||||||||||||||||||||||   || ||||
Db   2453 GCATACTGCTATATAGTTCGTCTCTC------CTTATCTCCCTCTCCTAACCTTCCTCT 2506
Qy   2735 TTCTTCTCCCATCCCATTCTTTTGAATTCCTCATCCCTCCAATCCGTATCTAC 2794
          ||| ||||||||||||
Db   2507 GTCTTCTTTTACCCAAAATTCTTTGAGTCCCGTCCTTCGTTCTTTTCTGCTCCCCCGTATCTAT 2566
Qy   2795 GCACC---CCCCCCCCCCAGGCAAAGCAGTGCTCTGAGTATCACATCACACAAAAGGAA 2851
          |||||   |||||  ||||  |  ||
Db   2567 GCACCCCTCCCCTCCCCCTCCCCCCAGGCAAAGCAGTGCTCTGAGTATCATCACACACAAGGAA 2626
Qy   2852 CAAAAGCGGAACACACAACCACACAGACCAGCCAGCTCAACTACACTTGGTACTCAAAGAACAAGAGT 2911
          |||||||||| | ||     ||||||||||
Db   2627 CAAATGCGAACACCACACAGACCAGCCAGCTCCAGCTTACTCACTCGGTTACTCACTCAAGCGAACAAGAGT 2686
Qy   2912 CAATGGTACTTGTCCTAGCGTTTTGGAAGAGAAAACAGGAAACCCACCCAGCGAAATAAACCAGC 2971
          ||||||||||||||| ||
Db   2687 CGGCGGTACTTGTCCTAGC-ATTCGGAAGAGGAAAGCAGGAGATCCCAGCAAATAAACCAGC 2745
Qy   2972 CAACCAAAGACAAAGAAAGAAAATTCCCACAATGAAAGAAATGTATTTGTCTTTTGCATTTGG 3031
          ||||| | |||||| |  ||
Db   2746 CAGCCAAAACAAAG-AAAAATCCCACAAGAAAGAAATGA----CTCTCTCCCATCTTGG 2800
Qy   3032 TGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTCTTTCTTTTAAAAAAAATGTCGAGG 3091
          |||||||||||||||||||
Db   2801 TATACAAGCCATCAACATTTAGTGAAATTCTTTCTTTCTTTAAAAATTTAAAGTGCAGGA 2860
Qy   3092 AAAGTAGAAATTTACCAAGGTTGTTGGCCCAGGCGTTAAATTCACAGATTTTTTTAACG 3151
          || |
Db   2861 AATAGCAATTTATACGAGGTTGTTGGCCCAGGCGTTAAATTTACAGATTTTTTTTAAA 2920
Qy   3152 AGAAAACACACAGA--AGAAGCTACCTCAGGTGTTTT 3188
          ||   |||||||||  ||
Db   2921 TGAGAAAAACACACACAACCACGAGCTACCTCCAGGTGTTTT 2960
```

```
RESULT 4
BQ691878
LOCUS       BQ691878                874 bp    mRNA     linear     EST 15-JUL-2002
DEFINITION  AGENCOURT_8034698 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6207026
            5', mRNA sequence.
ACCESSION   BQ691878
```

```
Db    421 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
          GCAAATCGTCTCTGTACATGTGTACATATTAGAAAGGGAAGCATGTTAAGATATGTGGC 480
Qy   2304 CTGTGGGTACACAGGGTGCCTGCAGCCGGTAATAATTAGAAAGGGAAGCATGTTAAGATATGTGGC 2363

Db    481 CTGTGGGTACACAGGGTGCCTGCAGCCGGTAATAATTA... 540
Qy   2364 CTCAACTAACTCCAATTTTTAATCAATTATTAATTATTA... 2423

Db    541 CTCAACTAACTCCAATTTTTAATCAATTATTAATTATTA... 600
Qy   2424 GGCTTTTCTAGACTTTAAAGAATAAAGTCTTGGGAGTCT... 2483

Db    601 GGCTTTTCTAGACTTTAAAGAATAAAGTCTTGGGAGTCT... 660
Qy   2484 TTGAGGCCACCCGACAAAATTCACCCAGAGGGAAATCGT... 2543

Db    661 TTGAGGCCACCCGACAAAATTCACCCAGAGGGAAATCGT... 720
Qy   2544 AGTTCTGATCACCTGTATGTCAACAGAAGGGATACCGTC... 2603

Db    721 AGTTCTGGATCACCTGTATGTCAACAGAAGGGATACCGT... 780
Qy   2604 TGTCACTCCTCATGCCTGCTCAGCTCATCATACACCCAT... 2663

Db    781 TGTCACTCCTCATGCCTGCTCAGCTCATCATACACCCAT... 840
Qy   2664 A--CTGGTTTTTGCATACTG 2682

Db    841 AACTGGGTTTTTGCATACTG 861
```

BM453327
LOCUS   BM453327   905 bp   mRNA   linear EST 05-FEB-2002
DEFINITION AGENCOURT_6397801 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5527679
            5', mRNA sequence.
ACCESSION   BM453327
VERSION   BM453327.1   GI:18502367
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 905)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12203 row: e   column: 24
            High quality sequence stop: 568.

```
Query Match               24.6%;    Score 841;  DB 12;  Length 905;
Best Local Similarity   97.1%;    Pred. No. 3.9e-154;
Matches  878;    Conservative   0;  Mismatches  21;  Indels  5;  Gaps  2;

Qy  492 CGGGCATCAGTTTGAGAACTACTCCTTCAGATTCCTACATCCCGATGAAGAGGTGAG 551
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 CGGGCATCAGTTTGAGAACTACTCCTTCAGATTCCTACATCCCGATGAAGAGGTGAG 60

Qy  552 CTCCCCTTCGCCCCCTCAGCGAGCCCAGCCGGTGGGGACCACTCTTCCGGGAGCAAGGCCA 611
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CTCCCCTTCGCCCCCTCAGCGAGCCCAGCCGGTGGGGACCACTCTTCCGGGAGCAAGGCCA 120

Qy  612 CGCCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTCCCGCTGCGGATCCTGGTCCC 671
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CGCCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTCCCGCTGCGGATCCTGGTCCC 180

Qy  672 CACCCAGTTGTTGGTGCCATCATCCGAAAGGAGGGCTTGACCATAAAGAACATCACTAA 731
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CACCCAGTTGTTGGTGCCATCATCCGAAAGGAGGGCTTGACCATAAAGAACATCACTAA 240

Qy  732 GCAGACCCAGTCCCGGTAGATATCCATAGAAAAGAGAAACTCTGGAGCTGCAGAGAAGCC 791
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GCAGACCCAGTCCCGGTAGATATCCATAGAAAAGAGAAACTCTGGAGCTGCAGAGAAGCC 300

Qy  792 TGTCACCATCCATGCCACCCAGAGGGGACTTCTGAAGCATGCCGAAGAGATTCCTCTGAAAT 851
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 TGTCACCATCCATGCCACCCAGAGGGGACTTCTGAAGCATGCCGAAGAGATTCCTCTGAAAT 360

Qy  852 CATGCAGAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAATCTTGGC 911
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 CATGCAGAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAATCTTGGC 420

Qy  912 ACACAATGGCTTGGTTGGTGGAAGACTGATTGGGAAAGAGAGGCAGAAATTTGAAGAAAATTGA 971
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ACACAATGGCTTGGTTGGTGGAAGACTGATTGGGAAAGAGAGGCAGAAATTTGAAGAAAATTGA 480

Qy  972 ACATGAAACAGGACCAAGATAACAATCACATCTTTGCAGGATATACACACCC 1031
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 ACATGAAACAGGACCAAGATAACAATCACATCTTTGCAGGATATACACACCC 540

Qy 1032 GGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGAGATAGAGAT 1091
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 GGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGAGATAGAGAT 600

Qy 1092 TATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTAACCAACAAGCCAA 1151
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 TATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTAACCAACAAGCCAA 660

Qy 1152 TCTGATCCCAGGGTTGAACCTCAGCGCACTTGCCATCTTTCAACAGGACTGTCCGTGCT 1211
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 TCTGATCCCAGGGTTGAACCTCAGCGCACTTGCCATCTTTCAACAGGACTGTCCGTGCT 720

Qy 1212 ATCTCCACCAGCAGGGCCCCGGAGTCCCCCGCTGCCCCCTACCACCCCTTCACTAC 1271
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 ATCTCCACCAGCAGGGCCCCGGAGTCCCCCGCTGCCCCCTACCACCCCTTCACTAC 780

Qy 1272 CCACTCCGGATACTTCTCCGACCTGTACCCCATCACCAGTTT--GGCCCGGTTCCCCGCAT 1329
        |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db  781 CCACTCCGGATACTTCTCCGACCTGTACCCCATCACCAGTTTTGGGCCCGGTTTGGCCCGCAT 840

Qy 1330 CATCACTCTTATCCAGAGCAGGAGGATTGTG---AATCTCTTCATCCCAACCCAGGCGTGTG 1386
        |||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db  841 CATCACTCTTATCCACGAGCAGCAGGAGATTGTGAAATCTCTTCATCCCAANCCCAGGCGTG 900

Qy 1387 GGGCG 1390
        ||||
Db  901 TGGG 904
```

RESULT 6
BQ221568
LOCUS   BQ221568   851 bp   mRNA   linear EST 02-MAY-2002

ORIGIN

```
Db    421  CAAAAATAGCCATGGAGAAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTT  480

Qy    527  CCTACATCCGGATGAAGAGGTGAGCTCCCTTCGCCCCTCAGCCGTGGGG  586
                |||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  CCTACATCCGGATGAAGAGGTGAGCTCCCTTCGCCCCTCAGCCGTGGGG  540

Qy    587  ACCACTCTTCCGGGAGCAAGGCCACGCGCCCTGGGGCACTTCTCAGGCCAGACAGATTG  646
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  ACCACTCTTCCGGGAGCAAGGCCACGCGCCCTGGGGCACTTCTCAGGCCAGACAGATTG  600

Qy    647  ATTTCCCGCTCGGGATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGGAGG  706
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  ATTTCCCGCTCGGGATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGGAGG  660

Qy    707  GCTTGACCATAAAGAACATCACTAAGCAGACCAGTCCCGGGTAGATATCCATAGAAAAG  766
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  GCTTGACCATAAAGAACATCACTAAGCAGACCAGTCCCGGGTAGATATCCATAGAAAAG  720

Qy    767  AGAACTCTGGAGCTGCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTG  826
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  AGAACTCTGGAGCTGCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTG  780

Qy    827  AAGCATGCCGCATGATTCTTGAAATCATGCAGAAGAGGCAGATGAGACCAAAC-TAGCC  885
                |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    781  AAGCATGCCGCATGATTCCTGGAATCCTGGAATTCCTGCAGAAAGAGGCAGATGAGACCCAACTTAGCC  840

Qy    886  GAAGAGATTC  895
                ||||||||||
Db    841  GAAGAGATTC  850
```

RESULT 7
BM561242

LOCUS   BM561242   999 bp   mRNA   linear   EST 20-FEB-2002
DEFINITION AGENCOURT_6567353 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5739550
            5', mRNA sequence.
ACCESSION   BM561242
VERSION   BM561242.1 GI:11806337
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 999)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12753 row: a   column: 23
            High quality sequence stop: 679.
FEATURES   Location/Qualifiers
     source   1..999
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5739550"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_88"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site 1: NotI; Site 2: SalI; Cloned unidirectionally.
                     Primer: oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
ORIGIN

DEFINITION AGENCOURT_7559207 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6046477
            5', mRNA sequence.
ACCESSION   BQ221568
VERSION   BQ221568.1 GI:20402968
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 851)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13292 row: f   column: 14
            High quality sequence stop: 637.
FEATURES   Location/Qualifiers
     source   1..851
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6046477"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_72"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
                     Site 2: SalI; Cloned unidirectionally. Primer: oligo dT.
                     Average insert size 2 kb. Library constructed by Life
                     Technologies."
ORIGIN

Query Match   24.3%; Score 830; DB 13; Length 851;
Best Local Similarity   99.3%; Pred. No. 5.6e-152;
Matches 844; Conservative   0; Mismatches   0; Indels   5; Gaps   1;

Qy   47 GCGGGCTCTCGGGGAAGAGAGACCGGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCG 106
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GCGGGCTCTCGGGGAAGAGAGACCGGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCG 60

Qy   107 CCGTCACCGCCGACGACCTCCGGCAGCTCTTTGGGACAGGAGAAGCTGCCCCTGGCGGGAC 166
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCGTCACCGCCGACGACCTCCGGCAGCTCTTTGGGACAGGAGAAGCTGCCCCTGGCGGGAC 120

Qy   167 AGGTCCTGCTGAAGTCCGGCTACGCCTTCGTCGGACTACCCGACCAGAACTGGGCCATCC 226
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGGTCCTGCTGAAGTCCGGCTACGCCTTCGTCGTGGACTACCCGACCAGAACTGGGCCATCC 180

Qy   227 GCGCCATCGAGACCCTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATT 286
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GCGCCATCGAGACCCTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATT 240

Qy   287 ACTCAGTCTCTAAAAAGCTAAGGAGCAGAGAAATTCAGATTCGAAACATCCTCCTCACC 346
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ACTCAGTCTCTAAAAAGCTAAGGAGCAGAGAAATTCAGATTCGAAACATCCTCCTCACC 300

Qy   347 TGCAGTGGGAGGTGTTGGATGGACTTTGGCTCAATATGGACAGTGGAGAATGTGGAAC 406
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TGCAGTGGGAGGTGTTGGATGGACTTTGGCTCAATATGGACAGTGGAGAATGTGGAAC 360

Qy   407 AAGTCAACACAGACACAGAAACCGCCGGTTGTCAACGTCACATATGCAACAAGAGAAGAAG 466
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AAGTCAACACAGACACAGAAACCGCCGGTTGTCAACGTCACATATGCAACAAGAGAAGAAG 420

Qy   467 CAAAAATAGCCATGGAGAAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTT 526
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

KEYWORDS　EST.
SOURCE　Homo sapiens (human)
ORGANISM　Homo sapiens
　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
　Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE　1 (bases 1 to 832)
AUTHORS　Lauber,J., Bahr,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
　Fobo,G., Han,M. and Wiemann,S.
TITLE　EST (Lauber,J., Bahr,A., Mewes,H.W., Weil,B., Amid,C., et al.)
JOURNAL　Unpublished (2003)
COMMENT　Contact: MIPS
　MIPS
　Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
　This is the 5' sequence of the clone insert
　Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
　Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
　sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
　consortium of the German Genome Project.
　No s1 sequence available.
　This clone (DKFZp781G0234) is available at the RZPD in Berlin.
　Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
　Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES　Location/Qualifiers
source　1..832
　/organism="Homo sapiens"
　/mol_type="mRNA"
　/db_xref="taxon:9606"
　/clone="DKFZp781G0234"
　/dev_stage="adult"
　/lab_host="DH10B"
　/clone_lib="781 (synonym: hlcc4)"
　/note="Vector: pSport1_sfi; Site_1: SfiIA; Site_2: SfiIB;
　cDNA-collection"
ORIGIN

Query Match　23.7%;　Score 809; DB 13; Length 832;
Best Local Similarity　99.9%;　Pred. No. 7.3e-148;
Matches 820;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　1;

Qy　1732 GAGGAAGTGATCGTCAGAGATTATCGGCACTTCCTTGCTAGCCAGACTGCACAGCGCAAG 1791
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　13 GAGGAAGTGATCGTCAGAGATTATCGGCACTTCCTTTGCTAGCCAGACTGCACAGCGCAAG 72

Qy　1792 ATCAGGGAAATTGTACAACAGTGAAGCAGCAGCAGCAGCAGCAGCAGAAATACCCTCAGC 1851
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　73 ATCAGGGAAATTGTACAACAGTGAAGCAGCAGCAGCAGCAGCAGCAGAAATACCCTCAGC 132

Qy　1852 TCACAGGCGCAGCAAGTGAGGCTCCCACAGCAGCACCAGCAAAACAACGGATGAATGAGTAGCCC 1911
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　 133 TCACAGGCGCAGCAAGTGAGGCTCCCACAGCAGCACCAGCAAAACAACGGATGAATGAGTAGCCC 192

Qy　1912 TTCCAACCTGACAGAATGAGACCATCTGAGGAAT 1971
　　　　　||||||||||||||||||||||||||||||||||
Db　 193 TTCCAACCTGACAGAATGAGACCATCTGAGGAAT 252

Qy　1972 CCATCTGAGGAATGAGAAGTCTGCGGAGGGCGGCCAGGGACTCTGCGGAGGCCCTGAGAAC 2031
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　 253 CCATCTGAGGAATGAGAAGTCTGCGGAGGGCGGCCAGGGACTCTGCGGAGGCCCTGAGAAC 312

Qy　2032 CCCAGGGGGGCCGGAGGAGGGGGGGGGCTTCAGGGCTTCAGGGAGTCAGGTTCAGCCCG 2091
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　 313 CCCAGGGGGGCCGGAGGAGGGGGGGGGCTTCAGGGCTTCAGGGAGTCAGGTTCAGCCCG 372

Qy　2092 CCTCCCGCCCCCCAGGGCTTCTCGCAGGCTATCCCTTTAGTTGAACTAAACATAGGTGAACGTGTTCA 2151
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　 373 CCTCCCGCCCCCCAGGGCTTCTCGCAGGCTATCCCTTTAGTTGAACTAAACATAGGTGAACGTGTTCA 432

Qy　2152 TCTTCCTGAACTCCCACGACGCTATCCCTTTTTAGTTGAACTAAACATAGGTGAACGTGTTCA 2211
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　 433 TCTTCCTGAACTCCCACGACGCTATCCCTTTTTGTCGTGCAAATGCTGTCTGTGTACATGTGTGTACA 492

Qy　2212 AAGCCAAGCAAAATGCACACCCTTTTCTGTGCGCAAATGCT 2271

Query Match　24.3%;　Score 829.8; DB 12; Length 999;
Best Local Similarity　97.5%;　Pred. No. 6e-152;
Matches 873;　Conservative　0;　Mismatches　19;　Indels　3;　Gaps　3;

Qy　1265 TCACTCACCACTCCGGATACTCTCCAGCCTGTACCCCATCACCAGTTGGCCGTTCC 1324
　　　　　||||||||||||| |||||||||| ||||||||||||||||||||||||||||||
Db　 108 TTAACACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTGGCCGTTCC 167

Qy　1325 CGCATCATCATCTCTTATCCAGAGCAGAGAGATTGTGAATCTCTTCATCCCAACCCAGGCTG 1384
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　 168 CGCATCATCATCTCTTATCCAGAGCAGAGAGGTTGTGAATCTCTTCATCCCAACCCAGGCTG 227

Qy　1385 TGGGGCCCATCATCGGAGAAGGGGGCCACACATCAAACAGCTGCGGAGATTCGCCGGAG 1444
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　 228 TGGGGCCCATCATCGGAGAAGGGGGCCACACATCAAACAGCTGGCGGAGATTCGCCGGAG 287

Qy　1445 CCTCTATCAAGATNTCCCCTGGGAGGCCCAGACGTCAGCGAAAGGATGGTCATCATCA 1504
　　　　　|||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db　 288 CCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCATCA 347

Qy　1505 CCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGACGGATCTTTGGGAAACTGAAAGAGG 1564
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　 348 CCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGACGGATCTTTGGGAAACTGAAAGAGG 407

Qy　1565 AAAACTTCTTAACCCAAAGAAGAAAGTGAAGCTGGAAGCGCATATCAGAGTGCCCTCTT 1624
　　　　　|||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db　 408 AAAACTTCTTTAACCCAAGAAGAAGAAAGTGAAGCTGGAAGCGCATATCAGAGTGCCCTCTT 467

Qy　1625 CCACAGCTGGCGCGGGTGATTGGCAAAGTGCAAGACCGTGAACGAACTGCAGAACTTAA 1684
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　 468 CCACAGCTGGCGCGGGTGATTGGCAAAGTGCAAGACCGTGAACGAACTGCAGAACTTAA 527

Qy　1685 CCAGTGCAGAAGTCATCGTGCCTCGTGACCAAAACGCCAGATGAAAATGAGGAAGTGATCG 1744
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　 528 CCAGTGCAGAAGTCATCGTGCCTCGTGACCAAAACGCCAGATGAAAATGAGGAAGTGATCG 587

Qy　1745 TCAGAATTATCGGGCACTTCCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAATTG 1804
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　 588 TCAGAATTATCGGGCACTTCCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAATTG 647

Qy　1805 TACAACAGTGAAGCAGCAGCAGCAGCAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAGCA 1864
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　 648 TACAACAGTGAAGCAGCAGCAGCAGCAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAGCA 707

Qy　1865 AGTGAGGCTCCCACAGCAGCACCAGCAGCAGATGGAGTGTAGCCCTTCCAACACCTGA 1924
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　 708 AGTGAGGCTCCCACAGCAGCACCAGCAGCAGATGGAGTGTAGCCCTTCCAACACCTGA 767

Qy　1925 CAGAATGAGACCAAAGCAGCAGCAAACGCAGCAGATCGGGAGACAAAGACCATCTGAGGAAT 1984
　　　　　|||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db　 768 CAGAATGAGACCAAAGCAGCAGCAAACGCAGCAGATCGGGAGACAAAGA-CATCTGAGGAAT 826

Qy　1985 GAGAAGTCTGCGGAGGGCGGCCAGGGACTCTGCCGGAGGGCCCAGGGGCCGAG 2044
　　　　　|||||||||||||||||||||||||||||||||||| ||||||||||||||
Db　 827 GAGAAGTCTGCGGAGGGCGGCCAGGGACTCTGCCGGAGAA-CCCAGGGGCCCAG 885

Qy　2045 GAGGGGGGGGAAGGTCAGGTTGCAGAACCACCACCACCACCACTCTCCTGA 2104
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||
Db　 886 GAGGGGCCGGGGAAAGTCACCCCAGGTTGCCAGAGACCCAGAACCCAGCCCCGCCTCCCGCCTCCCGGATCTCCCTGA 945

Qy　2105 AGGGGCTTCTGCGAGGCTTCAGCCATCCACATTCGGAACTCTCCTGA 2159
　　　　　|||||||||||||||||||| |||||
Db　 946 AAGGCTTCCTGCAGGGNTCAGCGGGNTCAGCCATCCATCCATCATTCA-CATCCATTCGGAACTCTCCTGA 999

RESULT 8
BX644668
LOCUS　　　BX644668　　832 bp　　mRNA　　linear　　EST 04-SEP-2003
DEFINITION　DKFZp781G0234_r1 781 (synonym: hlcc4) Homo sapiens cDNA clone
　　　　　　　DKFZp781G0234 5', mRNA sequence.
ACCESSION　BX644668
VERSION　　BX644668.1　GI:34479001

```
Db         493 AAGCCAAGCAAAATGCACACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTACA 552
Qy        2272 TATTAGAAAGGGAAGATGTTAAGATATGTGGCCTGGGTTACACAGGGTGCCTGCAGCG 2331
                    |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         553 TATTAGAAAGGGAAGATGTTAAGATATGTGGCCTGGGTTACACAGGGTGCCTGCAGCG 612
Qy        2332 GTAATATATTTTAGAAATAATAATATCAAATAACTAACTCCAATTTTAATCAATT 2391
Db         613 GTAATATATTTTAGAAATAATAATATCAAATAACTAACTCCAATTTTAATCAATT 672
Qy        2392 ATTAATTTTTTTTCTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGT 2451
Db         673 ATTAATTTTTTTTCTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGT 732
Qy        2452 CTTTGGGAGGTCTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGCACAAAATTCACCCA 2511
Db         733 CTTTGGGAGGTCTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGCAC-AAATTCACCCA 791
Qy        2512 GAGGGAAATCTCGTCGGAAGGACACTCACGGCAGTTCTGGA 2552
Db         792 GAGGGAAATCTCGTCGGAAGGACACTCACGGCAGTTCTGGA 832
```

RESULT 9
BM806132
LOCUS   BM806132   1089 bp   mRNA   linear   EST 05-MAR-2002
DEFINITION AGENCOURT_6553922 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5555652
            5', mRNA sequence.
ACCESSION   BM806132
VERSION   BM806132.1 GI:19122955
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1089)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapps-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12276 row: c column: 13
            High quality sequence stop: 688.
FEATURES   Location/Qualifiers
     source   1..1089
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5555652"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_71"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.1 kb. "
ORIGIN

Query Match   23.5%; Score 803.4; DB 12; Length 1089;
Best Local Similarity 94.0%; Pred. No. 8.8e-147;
Matches 936; Conservative 0; Mismatches 42; Indels 18; Gaps 9;

```
Qy         201 CTACCCCGACCACGAACTGGGCCATCCCGCCATCGAGACACCCTCTCCGGGTAAAGTGGAATT 260
                    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 CTACCCCGACCACGAACTGGGCCATCCCGCCATCGAGACACCCTCTCCGGGTAAAGTGGAATT 60
Qy         261 GCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGGAGCCAGGAAAAT 320
```

```
Db          61 GCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGGAGCCAGGAAAAT 120
Qy         321 TCAGATTGAAACATCCCTCCTCACCTGCGAGTGGAGGTGTTGGATGGACTTTGGGTCA 380
Db         121 TCAGATTGAAACATCCCTCCTCACCTGCGAGTGGAGGTGTTGGATGGACTTTGGGTCA 180
Qy         381 ATATGGACAGTGGAGAATGTGGAACAAGTCAACACAGACAAGAACCGCCGTTGTCAA 440
Db         181 ATATGGACAGTGGAGAATGTGGAACAAGTCAACACAGACAAGAACCGCCGTTGTCAA 240
Qy         441 CGTCACATATGCAACAGAGAGAGAAGCAAAAATAGCCATGAGAGAAGAGAGCGGGCATCA 500
Db         241 CGTCACATATGCAACAGAGAGAGAAGCAAAAATAGCCATGAGAGAAGAGAGCGGGCATCA 300
Qy         501 GTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGATGAAGAGAGGTGAGCTCCCCTTC 560
Db         301 GTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGATGAAGAGAGGTGAGCTCCCCTTC 360
Qy         561 GCCCCCTCAGCGACCCCAGCCGCCCAGCGTGGGGACCACTCTTCCCGGTGCGGATCCTGGTCTGGCTGCGGATCCTGGTCCCCACCCAGTT 620
Db         361 GCCCCCTCAGCGACCCCAGCCGCCCAGCGTGGGGACCACTCTTCCCGGTGCGGATCCTGGTCCCCACCCAGTT 420
Qy         621 GGGCACTTCTCAGGCCAGACAGATTGATTCCCGCGTTGCGGATCCTGGTCCCCACCCAGTT 680
Db         421 GGGCACTTCTCAGGCCAGACAGATTGATTCCCGCGTTGCGGATCCTGGTCCCCACCCAGTT 480
Qy         681 TGTTGGTGCCATCATCGGAAGGAGGGCTTGACCATAAGAACATCACTAAGCAGACCCA 740
Db         481 TGTTGGTGCCATCATCGGAAGGAGGGCTTGACCATAAAGACATCACTAAGCAGACCCA 540
Qy         741 GTCCCGGGTAGATATCCATAGAAAAGAGAACTCGGAGCTGCAGAGAAGCCTGTCACCAT 800
Db         541 GTCCCGGGTAGATATCCATAGAAAAGAGAACTCGGAGCTGCAGAGAAGCCTGTCACCAT 600
Qy         801 CCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCCATGATTCTTGAAATCATGCAGAA 860
Db         601 CCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCCATGATTCTTGAAATCATGCAGAA 660
Qy         861 AGAGGGCAGATGAGAACCAAACTAGCCGAAGAGATTCCTCTGAAAATCTTGGCACACAATGG 920
Db         661 AGAGGGCAGATGAGAACCAAACTAGCCGAAGAGATTCCTCTGAAANATCTTGGCACACAATGG 720
Qy         921 --CTTGGTTGGAAGACTGATTGGAAAAGAAGAGGCAGAAATTGAAGAAAATTGAACATGAA 978
Db         721 GCTTGGTTGGGAAGACTGATTGGAAAAGAAGAGGCAGAAATTGAAGAAAATTGAACATGTA 780
Qy         979 ACA-GGGACCAAGATAACAATCTCATCTTT-GCAGGATTTGAGCATATACAACCCGG-AA 1035
Db         781 ACAGGGGACCAAGATAACCATCTCATCTTTTGGCAGGATTTGAGCATATACCACCCGGAA 840
Qy        1036 AGAACCATCACTG-TGAAGGGCACAG-TTGAGGGCCTGTGTCCAGTGCCATGCTGAGATAGAGATTA 1093
Db         841 AGAACCATCACTGTTGAAGGGCACAGTTTGAGGGCCTGTGTCCATGCCATGCTGAAGATAGAGAT 900
Qy        1094 TGAAGAAGCTGC------GTGAGGCCTTTGAAAATGATATGCTGG--CTGTTAACCAAC 1144
Db         901 TATGGAAGAAACCTGGCCCGAGGCCCTTTTGAAAATGAAAATGCTGGCTGTTTAACCAAC 960
Qy        1145 AAGCC--AATCTGATCCCCAGGGTTGAACCTCAGCCGC 1178
Db         961 AAACCCAATCTTGATCCCCAGGGTTGAACCTCCACGC 996
```

RESULT 10
BQ212500
LOCUS   BQ212500   836 bp   mRNA   linear   EST 02-MAY-2002
DEFINITION AGENCOURT_7675661 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6095809
            5', mRNA sequence.
ACCESSION   BQ212500
VERSION   BQ212500.1 GI:20392782
KEYWORDS   EST.
SOURCE   Homo sapiens (human)

```
Qy   2708  TTTATCTCTCCCTCCCTCCCTCCCTTCTTCTCCATTCTTTGAATTTCCTC  2767
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  TTTATCTCTCCCTCCCTCCCTCCCTTCTTCTCCATTCTTTGAATTTCCTC   600

Qy   2768  ATCCCTCCATCTCAATCCGTATCTACGCACCCCCCCCCAGGCAAAGCAGTGCTCT  2827
                 |||||||||||||||||||||||||||||| ||||||||||||||||||
Db    601  ATCCCTCCATCTCAATCCGTATCTACGCA-CCCCCCCCCAGGCAAAGCAGTGCTCT   659

Qy   2828  GAGTATCACATCACACAAAAGGAACAAAAGGCGAAACACACAACCAGCCTCAACTTACAC  2887
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    660  GAGTATCACATCACACAAAAGGAACAAAAGGCGAAACACACAACCAGCCTCAACTTACAC   719

Qy   2888  TTGGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTA-GCGTTTGGAAGAGAAA  2946
                 |||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    720  TTGGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTANNCGTTTGGAAGAGAAA   779

Qy   2947  ACAGGAACCACCAAACCAACCAATCAACCAACCAAGAAAAAATTCCACCATGAAA  3003
                 ||||||||||||||||||||||||||||||||||    |||  |||
Db    780  ACAGGAACCACCAAACCAACCAATCAACCAACCAAGAAAAAATTCCACCATGAAA   836
```

RESULT 11
BU156245
LOCUS   BU156245   894 bp   mRNA   linear   EST 03-SEP-2002
DEFINITION   AGENCOURT_7965128 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6162660
              5', mRNA sequence.
ACCESSION   BU156245
VERSION   BU156245.1   GI:22669777
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 894)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: ATCC/DCTD/DTP
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM13516 row: o column: 13
              High quality sequence stop: 666.
FEATURES   Location/Qualifiers
     source   1..894
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6162660"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_72"
                     /note="Organ: sKin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2 kb. Library constructed by Life
                     Technologies."

ORIGIN

Query Match   22.3%;   Score 759.8; DB 13; Length 894;
Best Local Similarity   95.5%;   Pred. No. 3.1e-138;
Matches 836;   Conservative   0;   Mismatches   32;   Indels   7;   Gaps   5;

```
Qy   2102  CCCAGGGCTTCTGCAGGCTTCAGCCATCCACTTCACCATTCACCATCCGGATCTCTCCTGAAC  2161
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     20  CCCTCCTCACCTGCAGGCTTCAGCCATCCACTTCACCATCCACCATCCGGATCTCTCCTGAAC   79

Qy   2162  TCCCACGACGCTATCCTTTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCA  2221
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 836)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: ATCC/DCTD/DTP
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM13368 row: n column: 02
              High quality sequence stop: 670.
FEATURES   Location/Qualifiers
     source   1..836
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6095809"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_72"
                     /note="Organ: sKin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2 kb. Library constructed by Life
                     Technologies."

ORIGIN

Query Match   23.4%;   Score 799.2; DB 13; Length 836;
Best Local Similarity   98.7%;   Pred. No. 6e-146;
Matches 826;   Conservative   0;   Mismatches   9;   Indels   2;   Gaps   2;

```
Qy   2168  GACGCTATCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAATGC  2227
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  GACGCTATCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAATGC   60

Qy   2228  ACACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAAGGGGAAGA  2287
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  ACACCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAAGGGGAAGA  120

Qy   2288  TGTTAAGATATGTGGCCTGTGGGTTACACAGAGGGTGCCTGCAGCCGGTAATATTTTAGAA  2347
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  TGTTAAGATATGTGGCCTGTGGGTTACACAGAGGGTGCCTGCAGCCGGTAATATTTTAGAA  180

Qy   2348  ATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTAATTATTAATTTTTTTCT  2407
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  ATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTAATTATTAATTTTTTTCT  240

Qy   2408  TTTTAAAGAGAAAAGCAGCAGCCTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTCAC  2467
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  TTTTAAAGAGAAAAGCAGCAGCCTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTCAC  300

Qy   2468  GGTGTAGAGAGAGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGGGAAATCTCGTCG  2527
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  GGTGTAGAGAGAGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGGGAAATCTCGTCG  360

Qy   2528  GAAGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAGAGGGATACCGCTC  2587
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  GAAGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAGAGGGATACCGCTC  420

Qy   2588  CTTGAAGAGGAAACTCTGTCACTCCTCATGCCTGCTGCTCAGCTCATCACCATTTCTCTTT  2647
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  CTTGAAGAGGAAACTCTGTCACTCCTGTCACTGCTGCTCAGCTCATACACCCATTTCTCTTT  480

Qy   2648  GCTTCACCAGGTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCTCTCTCTG  2707
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  GCTTCACCAGGTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCTCTCTCTG  540
```

Tissue Procurement: DCTD/DTP
cDNA Library Preparation:Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10576 row: m column: 04
High quality sequence stop: 773.

FEATURES   Location/Qualifiers
    source   1..781
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4707243"
                     /tissue_type="mammary adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_87"
                     /note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.383 kb Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."

ORIGIN

Query Match   22.1%;   Score 752.6;   DB 12;   Length 781;
Best Local Similarity 98.6%;   Pred. No. 8.1e-137;
Matches 770;   Conservative   0;   Mismatches   9;   Indels   2;   Gaps   1;

Qy   1821 GCAGGAGCAGAAATACCCTCAGGGAGTGCGCTCACAGCGCAGCAAGTGAGGCTCCCACAG 1880
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GCAGGAGCAGAAATACCCTCAGGGAGTGCGCTCACAGCGCAGCAAGTGAGGCTCCCACAG 60

Qy   1881 GCACCAGCAAAACAACGGATGAATGAGACCAAAC 1940
            ||||||||||||||||||||||||||||||||||
Db   61 GCACCAGCAAAACAACGGATGAATGAGACCAAAC 120

Qy   1941 GCAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAGGAATGAGAAGTCTGCGGAGG 2000
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GCAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAGGAATGAGAAGTCTGCGGAGG 180

Qy   2001 CGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCCAG 2060
            ||||||||||||||||||||||||||||||||||||
Db   181 CGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCCAG 240

Qy   2061 CAGCCAGGTTTGCCAGAACCACCAGCCCCCGCCTCCCGCCTCCGGCTTTCTGCAGGCT 2120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CAGCCAGGTTTGCCAGAACCACCAGCCCCCGCCTCCCGCCTCCGGCTTTCTGCAGGCT 300

Qy   2121 TCAGCCATCCACTTCCACCATCCACTCGGATCTCTCCTGAACTCCCACGCGCTATCCCTT 2180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TCAGCCATCCACTTCCACCATCCACTCGGATCTCTCCTGAACTCCCACGCGCTATCCCTT 360

Qy   2181 TTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAGCAAAATGCACACCCTTTTTCT 2240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAGCAAAATGCACACCCTTTTTCT 420

Qy   2241 GTGGCAAATCGTCTCTGTACATGTGTACATATTGAAGGAGATGTTAAGATGT 2300
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GTGGCAAATCGTCTCTGTACATGTGTACATATTAGAAGGAGATGTTAAGATGT 480

Qy   2301 GGCCTGTGGGTTACACAGGGTGCCTGCAGGGTAATATATTTAGAAATAATAATATCAA 2360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GGCCTGTGGGTTACACAGGGTGCCTGCAGGGTAATATATTTAGAAATAATAATATCAA 540

Qy   2361 TAACTCAACTAACTCCAATTTTTAATCAATTATTAATTTTTTTTTTTTTTAAAGAGAAA 2420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TAACTCAACTAACTCCAATTTTTAATCAATTTTTAATCAATTATTAATTTTTTTTTTTTAAAGAGAAA 600

Qy   2421 GCAGGCTTTTCTAGACTTTAAAGAATACAGTCTTTGGGAGGTCTCACGGTGTAGAGAGGA 2480
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GCAGGCTTTTCTAGACTTTAAAGAATACAGTCTTTGGGAGGTCTCACGGTGTAGAGAGGA 660

Db   80 TCCCACGACGCTATCCCTTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCA 139

Qy   2222 AAATGCACACCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATATTAGAAAG 2281
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   140 AAATGCACACCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATATTAGAAAG 199

Qy   2282 GGAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCCGTAATATATT 2341
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   200 GGAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCCGTAATATATT 259

Qy   2342 TTAGAAATAATATATCAAAATAACTAACTAACTCCAATTTTTAATCAATTATTAATTTTT 2401
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   260 TTAGAAATAATATATCAAAATAACTCAACTAACTCCAATTTTTAATCAATTATTAATTTTT 319

Qy   2402 TTTTCTTTTTAAAGAGAAAGGAGCAGCGTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGG 2461
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   320 TTTTCTTTTTAAAGAGAAAGGAGCAGCGTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGG 379

Qy   2462 TCTCACGGTGTAGAGGAGCGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGAGGGAAATC 2521
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   380 TCTCACGGTGTAGAGGAGCGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGAGGGAAATC 439

Qy   2522 TCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCCTGTGTATGTCAACAGAGAGGGATAC 2581
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   440 TCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCCTGTGTATGTCAACAGAGAGGGATAC 499

Qy   2582 CGTCTCCTTGAAGAGGAAACTCTGTCACTCCTGTCTAGCTTCCTCATACACCCATTT 2641
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   500 CGTCTCCTTGAAGAGGAAACTCTGTCACTCCTGTCTAGCTTCCTCATACACCCATTT 559

Qy   2642 CTCTTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTC 2701
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   560 CTCTTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTC 619

Qy   2702 TCTCGTTTATCTCCCCTCCCCTCCCCTCCCCTTCTTCCATCTCA-TTCTTTTTGAA 2760
            |||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   620 TCTCGTTTATCTCCCCTCCCCTCCCCTCCCCTTCTTCCATCTCATTCTTTTTGAA 679

Qy   2761 TTTCCTCATCCCTCCATCTCAATCCGTATCCGACCACCCCCCCCCAGGCAAAGCA 2820
            |||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   680 TTTCCTCATCCCTCCATCTCAATCCGTATCCGTA--CCCCCCCCCAGGCAAAGCA 737

Qy   2821 GTGCTCTGAGTATCACATCACACAAAAAGGAACAAAAGCGAAACACACAGCCAGCCTCAA 2880
            |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   738 GTGCTCTGAGTATCACATCCCCAAAAGGACAAAAGCGAAACACACCAGCCAGCCTCAA 797

Qy   2881 C-TTACACTTGGTTACTCAAAG-AACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGG- 2937
            | ||||||||||||||||||| ||||| |||| ||||||||||||||||||||||
Db   798 CTTTACACTTGGTTACTCAAAGAAACAAACAAAGCCAGGTAACTTGGCCTAACCGTTTGGG 857

Qy   2938 -AAGAGGAAAACAGGAACCACCAAACCAACCAAT 2971
            |||||||||||||||||||||||||||||||||
Db   858 AAGAAGGAAAAACGGAAACCCCCAAACCCCCAAAACCCAACCAT 892

RESULT 12
BG575889

LOCUS   BG575889   781 bp   mRNA   linear   EST 10-APR-2001
DEFINITION 602598315F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4707243 5',
            mRNA sequence.
ACCESSION   BG575889
VERSION   BG575889.1 GI:13583542
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 781)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov

```
Qy   305  TAAGGAGCAGGAAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGGAGGTGTTGG  364
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   208  TAAGGAGCAGGAAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGGAGGTGTTGG  267

Qy   365  ATGGACTTTGGCTCAATATGGACAGTGGAGAATGTGGAACAAGTCAACAGACAGACACAG  424
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   268  ATGGACTTTGGCTCAATATGGACAGTGGAGAATGTGGAACAAGTCAACAGACAGACAG   327

Qy   425  AAACCGCCGTGTCAACGTCACATATGCAACAAGAGAAGCAAAAATAGCCATGGAGA       484
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   328  AAACCGCCGTGTCAACGTCACATATGCAACAAGAGAAGCAAAAATAGCCATGGAGA      387

Qy   485  AGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGATGAAG  544
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   388  AGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGATGAAG  447

Qy   545  AGGTGAGCTCCCCTTCGCCCCCTCAGCGAGCCCAGCGTGGGGACCCACTCTTCCCGGGAGC  604
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   448  AGGTGAGCTCCCCTTCGCCCCCTCAGCGAGCCCAGCGTGGGGACCCACTCTTCCCGGGAGC  507

Qy   605  AAGGCCACGCCCCTGGGGCACTTCTCAGCGCCAGACAGATTGATTTCCCGCTGCGGATCC  664
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   508  AAGGCCACGCCCCTGGGGCACTTCTCAGCGCCAGACAGATTGATTTCCCGCTGCGGATCC  567

Qy   665  TGGTCCCCACCCAGTTGTTGGTGCCATCATCCGAAAAGGAGGGCTTGACCATAAAGAACA  724
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   568  TGGTCCCCACCCAGTTGTTGGTGCCATCATCCGAAAAGGAGGGCTTGACCATAAAGAACA  627

Qy   725  TCACTAAGCAGACCAGTCCCGGGTAGATATCCATAGAAAAGAACTCTGGAGCTCGCAG    784
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   628  TCACTAAGCAGACCAGTCCCGGGTAGATATCCATAGAAAAGAACTCTGGAGCTCGCAG   687

Qy   785  AGAAGCCGTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATGATTC  844
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   688  AGAAGCCGTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATGATTC  747

Qy   845  TTGAAATCATGCA-GAAAGAGGCAGGAGATGAGACCAAAACTAG-CCGAAGAGATT-CCTCTGA  901
          |||||||||||| |||||||||||||||||||||||||||||| ||||||||||| |||||||
Db   748  TTGAAATCATGCAGGAAAGAGGCAGGAGATGAGACCAAAACTAGCCCGAAGAGATTCCCTCTGA  807

Qy   902  AAATCTTGG  910
          |||||||||
Db   808  AAATCTTGG  816

RESULT 14
CF593505
LOCUS           CF593505               824 bp    mRNA       linear    EST 26-SEP-2003
DEFINITION      AGENCOURT_15622249 NIH_MGC_147 Homo sapiens cDNA clone
                IMAGE:30531076 5', mRNA sequence.
ACCESSION       CF593505
VERSION         CF593505.1  GI:36347108
KEYWORDS        EST.
SOURCE          Homo sapiens (human)
ORGANISM        Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 824)
AUTHORS         NIH-MGC http://mgc.nci.nih.gov/.
TITLE           National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL         Unpublished (1999)
COMMENT         Contact: Daniela S. Gerhard, Ph.D.
                Office of Cancer Genomics
                National Cancer Institute / NIH
                Bldg. 31 Rm10A07 Bethesda, MD 20892
                Email: cgapbs-r@mail.nih.gov
                Tissue Procurement: Dr. Stefan Hansson
                cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
                and advice from Piero Carninci (RIKEN)
                cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Agencourt Bioscience Corporation
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
```

```
Qy  2481  GCTTTGAGGCCACCCGACACAAAATTCACCCAGAGGGAAATCTCGTCCGGAGGACACTCAC  2540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  GCTTTGAGGCCACCCGACACAAAATTCACCCAGAGGGAAATCTCGTCCGGAGGACACTCAC   720

Qy  2541  --GCAGTTCTGGATCACCTGTGTATGTCAACAGAAGGGATACCGTCTCCTTGAAGAGGA  2598
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721  CGGCAGTTTCTGGATCACCTGTGTATGTCAACAGAAGGGATACAGTCTCCTTGAAGAGGA   780

Qy  2599  A  2599
          |
Db   781  A  781

RESULT 13
CD243351
LOCUS           CD243351               940 bp    mRNA       linear    EST 22-MAY-2003
DEFINITION      AGENCOURT_14121251 NIH_MGC_180 Homo sapiens cDNA clone
                IMAGE:30383283 5', mRNA sequence.
ACCESSION       CD243351
VERSION         CD243351.1  GI:31003815
KEYWORDS        EST.
SOURCE          Homo sapiens (human)
ORGANISM        Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 940)
AUTHORS         NIH-MGC http://mgc.nci.nih.gov/.
TITLE           National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL         Unpublished (1999)
COMMENT         Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-r@mail.nih.gov
                Tissue Procurement: Dr. Michael Brownstein
                cDNA Library Preparation: Invitrogen Corp
                cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Agencourt Bioscience Corporation
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: NDAM446 row: o column: 04
                High quality sequence start: 18
                High quality sequence stop: 695.
FEATURES             Location/Qualifiers
     source          1..940
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30383283"
                     /lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
                     /clone_lib="NIH_MGC_180"
                     /note="Organ: Testis; Vector: pCMV-SPORT6.1; Site_1: NotI;
                     Site_2: EcoRV (destroyed); Library is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed upon
                     cloning). Average insert size 1.68 kb. Library was
                     constructed by (Invitrogen). Note: this is a NIH_MGC
                     Library."

ORIGIN
Query Match       22.0%;   Score 751.4;  DB 14;  Length 940;
Best Local Similarity  99.5%;  Pred. No. 1.4e-136;
Matches  785;  Conservative    0; Mismatches   1; Indels   1;  Gaps   3;

Qy   125  TCCGGCAGCTCTTTGGGACAGGAAGCTGCCCTGCGGGGACCAGGTCCTGCTGAAGTCCG  184
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    28  TCCGGCAGCTCTTTGGGGACAAGGAAGCTGCCCTGCGGGGACCAGGTCCTGCTGAAGTCCG   87

Qy   185  GCTACGCGTTCGTGGACTACCCCGACCAGACTGGGGCACTCGCGGCCATCCGCGCCATCCGAGACCCTCT  244
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    88  GCTACGCGTTCGTGGACTACCCCGACCAGACTGGGGCACTCGCGGCCATCCGCGCCGAGACCCCTCT  147

Qy   245  CGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGC  304
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   148  CGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGC  207
```

```
Db    698 ATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCANTAAG 757
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Qy    721 AACATCACTAAGCACAGACCCAGTCC-GGTAGATATCCATA-GAAAAGAGAACTCTGG 776
          |||||||||||||||||||||||||| |||||||||||||| ||||||||||||||||
Db    758 AACATCACTAAGCACAGACCCAGTCCCGGGGTAGATATCCTTAGGAAAAGAGAACTCTGG 815
```

RESULT 15
BG748346
LOCUS   BG748346   602705902F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4842553 5',
               mRNA sequence.
DEFINITION   602705902F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4842553 5',
               mRNA sequence.
ACCESSION   BG748346
VERSION   BG748346.1 GI:14058999
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 889)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: ATCC
               cDNA Library Preparation: Ling Hong/Rubin Laboratory
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLCM1677 row: c column: 02
               High quality sequence stop: 773.
FEATURES   Location/Qualifiers
     source   1..889
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4842553"
                     /tissue_type="normal pigmented retinal epithelium"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC 43"
                     /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library. |"

ORIGIN

Query Match   21.9%;   Score 746; DB 12;   Length 889;
Best Local Similarity 95.5%; Pred. No. 1.6e-135;
Matches 802; Conservative 0; Mismatches 30;   Indels   8; Gaps   3;

Qy   72 GATGAACAAGCTTACATCGGGAACCTGAGCCCCGCCCGTCACCGCGAGACCTCCGGCA 131
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 GATGAACAAGCTTTACATCGGGAACCTGAGCCCCGCCCGTCACCGCGAGACCTCCGGCA 61

Qy   132 GCTCTTTGGGGACAGGAAGCTGCCCCTGGCGCGGGACCAGGTCCTGCGGAGTCCGGCTACGC 191
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 GCTCTTTGGGGACAGGAAGCTGCCCCTGGCGCGGGACCAGGTCCTGCGTGAAGTCCGGCTACGC 121

Qy   192 CTTCGTGGACTACCCCGACCAGAATGGGCCATCCGCGCCGCCATCGAGACCTCTCGGGTA 251
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 CTTCGTGGACTACCCCGACCAGAACTGGCCATCCGCGCCGCCATCGAGACCTCTCGGGTAA 181

Qy   252 AGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTTAAAAAGCTAAGGAG 311
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   182 AGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTTAAAAAGCTAAGGAG 241
```

(Left column, rotated text)

```
http://image.llnl.gov
Plate: NDAM621 row: e  column: 05
High quality sequence stop: 682.
FEATURES        Location/Qualifiers
     source     1..824
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:30531076"
                /tissue_type="Human Placenta"
                /lab_host="DH10B TonA"
                /clone_lib="NIH_MGC 147"
                /note="Organ: placenta; Vector: pBluescriptR; Site_1:
                aII-XhoI; Site_2: BamH; Oligo-dT primed using primer
                5'-TTTTTTTTTTTTTTVN-3', size-selected for average
                insert size 2.3 kb and normalized to ROT 5. This is a
                primary library enriched for full-length clones and
                constructed using the Cap-trapper method (Carninci, in
                preparation). Library constructed by M. Brownstein
                (NIMH/NHGRI, National Institutes of Health). Note: This is
                a NIH_MGC library."

ORIGIN

Query Match    22.0%;   Score 749.8;  DB 14;   Length 824;
Best Local Similarity 99.4%;  Pred. No. 2.9e-136;
Matches  773;  Conservative  0;  Mismatches  3;   Indels   2;  Gaps   2;

Qy      1 GGCAGCGGAGGAGGCGAGGAGGCGCGGGTACCCGGGCCGGGGAGCCGCTCCTCGGGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     38 GGCAGCGGAGGAGGCGAGGAGGCGCGGGTACCCGGGCCGGGGAGCCGCCGGGGGCTCCTCGGGG 97

Qy     61 AAGAGACGGATGATGAACAAGCTTTACATCGGGAAACCTGAGCCCGCCCCGTCACCGCCGACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     98 AAGAGACGGATGATGAACAAGCTTTACATCGGGAAACCTGAGCCCGCCCCGTCACCGCCGACA 157

Qy    121 GACCTCCGGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGGCGCGGGACAGGTCCTGCTGAAG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    158 GACCTCCGGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGGCGCGGGACAGGTCCTGCTGAAG 217

Qy    181 TCCGGCTACGCCTTCGTGGACTACCCCGACCAGAACTGGCCATCCGCGCCCAGAGACC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    218 TCCGGCTACGCCTTCGTGGACTACCCCGACCAGAACTGGCCATCCGCGCCCAGAGACC 277

Qy    241 CTCTCGGGTAAAGTGGAATTGGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTTCTAAA 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    278 CTCTCGGGTAAAGTGGAATTGGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTTCTAAA 337

Qy    301 AAGCTAAGGAGCAGCAGAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGGAGGTG 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    338 AAGCTAAGGAGCAGCAGAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGGAGGTG 397

Qy    361 TTGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGAC 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    398 TTGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGAC 457

Qy    421 ACAGAAAACCGCCGTTGTCAACGTCACATATGGGACAGTCAACGTCAACAAAAATAGCCATG 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    458 ACAGAAAACCGCCGTTGTCAACGTCACATATGGGACAGTCAACGTCAACAAGAAGAAGCAAAAATAGCCATG 517

Qy    481 GAGAAGCTAAGCGGCATCGGGCATCAGTTTGAGAACTACTCTTCAAGATTTCCTACATCCCGGAT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    518 GAGAAGCTAAGCGGCATCGGGCATCAGTTTGAGAACTACTCTTCAAGATTTCCTACATCCCGGAT 577

Qy    541 GAAGAGGTGAGCTCCCCTTCGCCCCTTCAGCGAGCCCAGCGTGGGGAGCCACTCTTTCCCGG 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    578 GAAGAGGTGAGCTCCCCTTCGCCCCTTCAGCGAGCCCAGCGTGGGGAGACCACTCTTTCCCGG 637

Qy    601 GAGCAAGGCCACGCCGCCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGG 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    638 GAGCAAGGCCACGCCGCCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGG 697

Qy    661 ATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy   312 CAGGAAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGGAGGTGTTGGATGACT 371
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   242 CAGGAAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGGAGGTGTTGGATGACT 301

Qy   372 TTTGGCTCAATATGGGACAGTGGAACAAGTCAACAGAACAGAGAAACCGC 431
         |||||||||||| ||||||||||||||||||||||||||||||||||||
Db   302 TTTGGCTCAATATGGGACAGTGGAACAAGTCAACAGAACAGAGAAACCGC 361

Qy   432 CGTTGTCAACGTCACATATGCAACAAGAAGCAAAAATAGCCATGGAGAAGTAAG 491
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   362 CGTTGTCAACGTCACATATGCAACAAGAAGCAAAAATAGCCATGGAGAAGTAAG 421

Qy   492 CGGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGATGAGAGGTGAG 551
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   422 CGGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGATGAGAGGTGAG 481

Qy   552 CTCCCCTTCGCCCCCTCAGCGAGCCAGCGTGGGGACCACTCTTCCCGGGAGCCAAGGCCA 611
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   482 CTCCCCTTCGCCCCCTCAGCGAGCCAGCGTGGGGACCACTCTTCCCGGGAGCCAAGGCCA 541

Qy   612 CGCCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCCGATCCTGGTCCC 671
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   542 CGCCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCCGATCCTGGTCCC 601

Qy   672 CACCCAGTTTGTTGTTGGTGCCATCATCGGAAAGAGGAGGGCTTGACCATAAAGAACATCACTAA 731
         ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   602 CAACCAGTTTGTAGGTGCCATCATCGGAAAGAGGAGGGCTTGACCATAAAGAACATCACTAA 661

Qy   732 GCAGACCCAGTCCCCGGTAGATATCCATAGAAAAGAGAACTCTGGAGCTGCAGAGAAGCC 791
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   662 GCAGACCCAGTCCCCGGTAGATATCCATAGAAAAGAGAACTCTGGAGCTGCAGAGAAG-C 720

Qy   792 TGTCACCATCCATGCACCCCAGAGGGGACTTGAA--GCATGCCGCATGATGATTCTTGAA 849
         |||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db   721 TGTCACCATCCATGCACCCCAGAGGGGACTTCTTAAAGCATTGCCGCATGATTCTCTGA 780

Qy   850 ATCATGCAGAAAGAGGC-----AGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAA 904
         |||||||||||||||||     | |||||||||||||||||||||||| |||| |||||
Db   781 AATCCTGGCAGACAGACCCGCAGATTTGAGACCAAAATAGCCGGAGAAATCGTCTGTAAA 840
```

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2004 Compugen Ltd.

GenCore version 5.1.6

Run on: July 20, 2004, 21:15:35 ; Search time 12992 seconds
                                         (without alignments)
                                         11382.884 Million cell updates/sec

Title: US-09-270-437D-6
Perfect score: 3412
Sequence: 1 ggcagcggaggaggcgagga.........aaccttgaaaatgttattt 3412

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
             1: gb_ba:*
             2: gb_htg:*
             3: gb_in:*
             4: gb_om:*
             5: gb_ov:*
             6: gb_pat:*
             7: gb_ph:*
             8: gb_pl:*
             9: gb_pr:*
            10: gb_ro:*
            11: gb_sts:*
            12: gb_sy:*
            13: gb_un:*
            14: gb_vi:*
            15: em_ba:*
            16: em_fun:*
            17: em_hum:*
            18: em_in:*
            19: em_mu:*
            20: em_om:*
            21: em_or:*
            22: em_ov:*
            23: em_pat:*
            24: em_ph:*
            25: em_pl:*
            26: em_ro:*
            27: em_sts:*
            28: em_un:*
            29: em_vi:*
            30: em_htg_hum:*
            31: em_htg_inv:*
            32: em_htg_other:*
            33: em_htg_mus:*
            34: em_htg_pln:*
            35: em_htg_rod:*
            36: em_htg_man:*
            37: em_htg_vrt:*
            38: em_sy:*
            39: em_htgo_hum:*
            40: em_htgo_mus:*
            41: em_htgo_other:*

SUMMARIES

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3411	100.0	3412	6	AR171865	AR171865 Sequence
2	3411	100.0	3412	6	AR343074	AR343074 Sequence
3	3411	100.0	3412	6	BD209925	BD209925 Isolated
4	3383	99.2	3633	9	BC021290	BC021290 Homo sapi
5	3143	92.1	3283	6	AR171867	AR171867 Sequence
6	3143	92.1	3283	6	AR343076	AR343076 Sequence
7	3143	92.1	3283	6	BD209927	BD209927 Isolated
8	3114.8	91.3	3667	9	AF057352	AF057352 Homo sapi
9	2662	78.0	142971	9	AC020629	AC020629 Homo sapi
10	2644	77.5	91084	9	AL596177	AL596177 Human DNA
11	2642.4	77.4	182695	9	AC015706	AC015706 Homo sapi
12	2189.6	64.2	187226	9	AC104980	AC104980 Homo sapi
13	2188	64.1	113201	9	AP004290	AP004290 Homo sapi
c	2165	63.5	3557	10	BC023758	BC023758 Mus muscu
15	1979.8	58.0	2010	9	AF117107	AF117107 Homo sapi
c	1627	47.7	98945	9	AC016961	AC016961 Homo sapi
17	1618	47.4	2640	9	HSM806243	BXS37583 Homo sapi
c	1613	47.3	208273	2	AC134942	AC134942 Homo sapi
19	1234.4	36.2	4953	10	BC054552	BC054552 Mus muscu
20	748.6	21.9	156398	2	AC094924	AC094924 Rattus no
21	743.8	21.8	2021	5	AF026527	AF026527 Gallus ga
22	715.4	21.0	2105	10	AF541940	AF541940 Rattus no
23	712	20.9	1740	6	BD275902	BD275902 COMPOUNDS
24	712	20.9	1740	6	AR220687	AR220687 Sequence
25	712	20.9	1740	6	AR281251	AR281251 Sequence
26	712	20.9	1740	6	AX365954	AX365954 Sequence
27	712	20.9	1743	6	AX366054	AX366054 Sequence
28	712	20.9	1743	6	AX366057	AX366057 Sequence
29	710.4	20.8	4115	9	HSU76705	U76705 Human putat
30	710.4	20.8	4159	6	AR171863	AR171863 Sequence
31	710.4	20.8	4159	6	AR343072	AR343072 Sequence
32	710.4	20.8	4159	6	BD209923	BD209923 Isolated
33	710.4	20.8	4171	9	AF117108	AF117108 Homo sapi
34	710.4	20.8	4181	6	BD275767	BD275767 COMPOUNDS
35	710.4	20.8	4181	6	AR220552	AR220552 Sequence
36	710.4	20.8	4181	6	AR255546	AR255546 Sequence
37	710.4	20.8	4181	6	AR281116	AR281116 Sequence
38	710.4	20.8	4181	6	AX333233	AX333233 Sequence
39	710.4	20.8	4181	6	AX365782	AX365782 Sequence
40	710.4	20.8	4181	9	HSU97188	U97188 Homo sapien
41	710.4	20.8	4602	6	AX397963	AX397963 Sequence
42	709.8	20.8	2697	5	AF064633	AF064633 Xenopus 1
43	709.8	20.8	3147	5	AF042353	AF042353 Xenopus 1
44	709.4	20.8	2031	5	BC057700	BC057700 Xenopus 1
45	709	20.8	1764	6	AX366035	AX366035 Sequence

ALIGNMENTS

RESULT 1
AR171865
LOCUS   AR171865   3412 bp   DNA   linear   PAT 17-DEC-2001
DEFINITION   Sequence 6 from patent US 6297364.
ACCESSION   AR171865
VERSION   AR171865.1 GI:17910815
KEYWORDS   .
SOURCE   Unknown.
ORGANISM   Unknown.
             Unclassified.
REFERENCE   1 (bases 1 to 3412)
AUTHORS   Chen,Y.-T., Gure,A., Tsang,S., Stockert,E., Jager,E., Alexander,K.
             and Old,L.J.
TITLE   Isolated nucleic acid molecule encoding cancer associated antigen,
             the antigen itself, and uses thereof

Pred. No. is the number of results predicted by chance to have a

JOURNAL   Patent: US 6297364-A 6 02-OCT-2001;
FEATURES
source   1..3412
                /organism="unknown"
                /mol_type="unassigned DNA"
ORIGIN

Query Match   100.0%;   Score 3411;   DB 6;   Length 3412;
Best Local Similarity   100.0%;   Pred. No. 0;
Matches 3412;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```
Qy      1 GGCAGCGGAGGAGGCGGAGGAGGCGCCGGGTACCGGCGGCGGGGCTCCGGGG     60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGCAGCGGAGGAGGCGGAGGAGGCGCCGGGTACCGGCGGCGGGGCTCCGGGG     60

Qy     61 AAGAGACGGATGATGAACAAGCTTTACATCGGAGAACCTGAGCCCGTCACCGCCGAC    120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAGAGACGGATGATGAACAAGCTTTACATCGGAGAACCTGAGCCCGTCACCGCCGAC    120

Qy    121 GACCTCCGGCAGCTCTTTGGGACAGGAGCTGCCCCTGGCGGGACAGTCCTGTGAAG    180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GACCTCCGGCAGCTCTTTGGGACAGGAGCTGCCCCTGGCGGGACAGTCCTGTGAAG    180

Qy    181 TCCGGCTACGCCTTCGTGACTACCCCGACCAGAACTGGGCCATCACCGGCCATCGAGACC    240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCCGGCTACGCCTTCGTGACTACCCCGACCAGAACTGGGCCATCACCGGCCATCGAGACC    240

Qy    241 CTCTCGGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA    300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTCTCGGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA    300

Qy    301 AAGCTTAAGGAGCAGGAAAATTCAGATTCAGATTCGAAACATCCTCCTCCAGTCGGTGGGAGGTG    360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AAGCTTAAGGAGCAGGAAAATTCAGATTCAGATTCGAAACATCCTCCTCCAGTCGGTGGGAGGTG    360

Qy    361 TTGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGAACAAGTCAACACAGAC    420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGAACAAGTCAACACAGAC    420

Qy    421 ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAGAGAAGCAAAAATAGCCATG    480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAGAGAAGCAAAAATAGCCATG    480

Qy    481 GAGGAAAACTTCTTTAACCCAAAGAGAGTGGCAAAGGTGGCAAAGGCCATATCAGAGTGCCC    540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GAGGAAAACTTCTTTAACCCAAAGAGAGTGGCAAAGGTGGCAAAGGCCATATCAGAGTGCCC    540

Qy    541 TCTTCCACAGTGGCGCGGGTGAGCTCCCCTTCGCCCCTCAGCCGAGCCACTCTTCCCGG    600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TCTTCCACAGTGGCGCGGGTGAGCTCCCCTTCGCCCCTCAGCCGAGCCACTCTTCCCGG    600

Qy    601 GAGGCAAGGCCACCGCCCCGGGGGACACTTGTGGTGCCATCATCGAGAAGATTGAATTGGG    660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GAGGCAAGGCCACCGCCCCGGGGGACACTTGTGGTGCCATCATCGAGAAGATTGAATTGGG    660

Qy    661 ATCCTGGTCCCCACCCAGTTGTTGTGGTCCATCATCGGAAAGGAGGGCTTGACCATAAAG    720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ATCCTGGTCCCCACCCAGTTGTTGTGGTCCATCATCGGAAAGGAGGGCTTGACCATAAAG    720

Qy    721 AACATCACTAAGCAGAGGGTAGATATCCCGGTAGATATCCATAGAAAAGAGAACTCTGAGCT    780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AACATCACTAAGCAGAGGGTAGATATCCCGGTAGATATCCATAGAAAAGAGAACTCTGAGCT    780

Qy    781 GCAGAGAAGCCTGTCACCATCCACCCCATGCCACCCATGGGACTTCTGAAGCATGCCGCCATG    840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GCAGAGAAGCCTGTCACCATCCACCCCATGCCACCCATGGGACTTCTGAAGCATGCCGCCATG    840

Qy    841 ATTCTTGAAATCATGCAGAAAGAGGAGGAATGAGATGAAAGAGAGGAATTCCTCTG    900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 ATTCTTGAAATCATGCAGAAAGAGGAGGAATGAGATGAAAGAGAGGAATTCCTCTG    900

Qy    901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTG    960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTG    960

Qy    961 AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTGAGC    1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTGAGC    1020

Qy   1021 ATATACACCCGAAAGAACCATCACTGTGAGGGACCAGTTGAGGGCCTGTCCCAGTGCT    1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ATATACACCCGAAAGAACCATCACTGTGAGGGACCAGTTGAGGGCCTGTCCCAGTGCT    1080

Qy   1081 GAGATAGAGATTATGAGAGAAGCTGCGTGAGGCCTTTGAAAATGATATGTCGGCTGTTAAC    1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GAGATAGAGATTATGAGAGAAGCTGCGTGAGGCCTTTGAAAATGATATGTCGGCTGTTAAC    1140

Qy   1141 CAACAAGCCAATCTGATCCGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGA    1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CAACAAGCCAATCTGATCCGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGA    1200

Qy   1201 CTGTCCGTGCTATCTCCACCAGCAGGGCCCCGGGAGCTCCCCCCGCGGAGTCCTGCTGAAG    1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CTGTCCGTGCTATCTCCACCAGCAGGGCCCCGGGAGCTCCCCCCGCGGAGTCCTGCTGAAG    1260

Qy   1261 CCCTTCACTACCCACTCCGGATACTTCTCCAGCAGGACCGTGTACCCCATCACCAGTTGGCCCG    1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CCCTTCACTACCCACTCCGGATACTTCTCCAGCAGGACCGTGTACCCCATCACCAGTTGGCCCG    1320

Qy   1321 TTCCCGCATCATCACTCTTATCCGGGAAGACCAGGAAGATTGTGAATTGTGAATCTCTTCATCCCAACCCAG    1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TTCCCGCATCATCACTCTTATCCGGGAAGACCAGGAAGATTGTGAATTGTGAATCTCTTCATCCCAACCCAG    1380

Qy   1381 GCTGTGTGGGCGCCATCATCGGGGAAGAAGAGGGGCACACATCAAACAGCTGGCCGAGATTCGCC    1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GCTGTGTGGGCGCCATCATCGGGGAAGAAGAGGGGCACACATCAAACAGCTGGCCGAGATTCGCC    1440

Qy   1441 GGAGCCTCTATCATCAAGATTGCCCCTCGCGGGAAGGCCCAGAACGTCAGCGAAGGATGGTCATC    1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GGAGCCTCTATCATCAAGATTGCCCCTCGCGGGAAGGCCCAGAACGTCAGCGAAGGATGGTCATC    1500

Qy   1501 ATCACCGGCCACCGGAGCCCAGTTCAAGGCCCCAGGGACGGATCTTTGGGAAACTGAAA    1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 ATCACCGGCCACCGGAGCCCAGTTCAAGGCCCCAGGGACGGATCTTTGGGAAACTGAAA    1560

Qy   1561 GAGGAAACTTCTTTAACCCAAAGAGAGTGAAGCTGGCAAAGGTGGCAAAGGCGCCCC    1620
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 GAGGAAACTTCTTTAACCCAAAGAGAGTGAAGCTGGCAAAGGTGGCAAAGGCGCCCC    1620

Qy   1621 TCTTCCACAGTGGCGGGTGATTGGCAAAGGTGGCAAAGGTGCAAGACCGTGAACGAAC    1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TCTTCCACAGTGGCGGGTGATTGGCAAAGGTGGCAAAGGTGCAAGACCGTGAACGAAC    1680

Qy   1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAGAGTGAAAATGAGGAAGTG    1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAGAGTGAAAATGAGGAAGTG    1740

Qy   1741 ATCGTCAGAATTATCGGGCACTTCTTTGCTTAGCCAGACTGCACAGCGCAAGATCAGGGAA    1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 ATCGTCAGAATTATCGGGCACTTCTTTGCTTAGCCAGACTGCACAGCGCAAGATCAGGGAA    1800

Qy   1801 ATTGTACAACAGGTGAAGCAGCAGAGCAGAAAATACCCTCAGGAGTCGGCCTCACACGCGC    1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 ATTGTACAACAGGTGAAGCAGCAGAGCAGAAAATACCCTCAGGAGTCGGCCTCACACGCGC    1860

Qy   1861 AGCAAGTGAGGCTCCCACCCAGGCACCAGCAGAAACAACGGATGAATTGTAGCCCTTCCAAAC    1920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 AGCAAGTGAGGCTCCCACCCAGGCACCAGCAGAAACAACGGATGAATTGTAGCCCTTCCAAAC    1920

Qy   1921 CTGACAGAATGAGACAGAGACAGCAGCAGGCAAACGCAGCCGATCGGGAGCAAACCATTCTGAG    1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CTGACAGAATGAGACAGAGACAGCAGCAGGCAAACGCAGCCGATCGGGAGCAAACCATTCTGAG    1980

Qy   1981 GAATGAGAAGTCTGCGGAGGCGGCCAGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGC    2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GAATGAGAAGTCTGCGGAGGCGGCCAGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGC    2040
```

```
Qy  3121 CAGGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAGAAGCTACCTCAG 3180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3121 CAGGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAGAAGCTACCTCAG 3180

Qy  3181 GTGTTTTACCTCAGCACCTTGCTCTGTGTTCCCTTAGAGATTTTGTAAAGCTGATAG 3240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3181 GTGTTTTACCTCAGCACCTTGCTCTGTGTTCCCTTAGAGATTTTGTAAAGCTGATAG 3240

Qy  3241 TTGGAGCATTTTTTATTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACT 3300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3241 TTGGAGCATTTTTTATTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACT 3300

Qy  3301 GCCAGCCTGGAGAAGGTGACAGTGCCAAGTGTCTTGCAACAGCTGTTCTGAATTGTCTTCCGCT 3360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3301 GCCAGCCTGGAGAAGGTGACAGTGCCAAGTGTCCAAGCTGTTCTGAATTGTCTTCCGCT 3360

Qy  3361 AGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3412
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3361 AGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3412
```

RESULT 2

```
LOCUS         AR343074                    3412 bp        DNA        linear    PAT 17-AUG-2003
DEFINITION    Sequence 6 from patent US 6576756.
ACCESSION     AR343074
VERSION       AR343074.1  GI:33738476
KEYWORDS      .
SOURCE        Unknown.
  ORGANISM    Unknown.
              Unclassified.
REFERENCE     1  (bases 1 to 3412)
  AUTHORS     Chen,Y.-T., Gure,A., Tsang,S., Stockert,E., Jager,E., Alexander,K.
              and Old,L.J.
  TITLE       Isolated nucleic acid molecule encoding cancer associated antigen,
              the antigen itself, and uses thereof
  JOURNAL     Patent: US 6576756-A 6 10-JUN-2003;
FEATURES      Location/Qualifiers
  source      1..3412
              /organism="unknown"
              /mol_type="genomic DNA"
ORIGIN
```

```
Query Match            100.0%;  Score 3411;  DB 6;  Length 3412;
Best Local Similarity  100.0%;  Pred. No. 0;
Matches 3412;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
Qy    1 GGCAGCGGAGGGGCGGGAGGAGGCGCGGGTACCGGGCGCCGGGGGAGGGCCGGGGAGGGCCTCTCGGGG 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GGCAGCGGAGGGGCGGGAGGAGGCGCGGGTACCGGGCGCCGGGGGAGGGCCGGGGAGGGCCTCTCGGGG 60

Qy   61 AAGAGACGGATGATGAACAAGCTTTACATCGGAACCTTGAGCCCGCCGTCACCGGCCGAC 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AAGAGACGGATGATGAACAAGCTTTACATCGGAACCTTGAGCCCGCCGTCACCGGCCGAC 120

Qy  121 GACTTCCGGCAGCTCTTTGGGACAGGAAGCTGCCCTGGGACAGGTCCTGCTGAAG 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GACTTCCGGCAGCTCTTTGGGACAGGAAGCTGCCCTGGGACAGGTCCTGCTGAAG 180

Qy  181 TCCGGCTACGCCTTCGTGGTCGTGGACTACCCCGGACTAGCCGGAAATCATGGAAGTTCAGTCTCTAA 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TCCGGCTACGCCTTCGTGGTCGTGGACTACCCCGGACTAGCCGGAAATCATGGAAGTTCAGTCTCTAA 240

Qy  241 CTCTCGGGTAAAGTGCATGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAA 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CTCTCGGGTAAAGTGCATGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAA 300

Qy  301 AAGCTAAGGAGCAGGAGAAAATTCAGATTCAGAATCGAAGTAGCCCTCCTCACCTGCTGGGAGGTG 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 AAGCTAAGGAGCAGGAGAAAATTCAGATTCAGAATCGAAGTAGCCCTCCTCACCTGCTGGGAGGTG 360

Qy  361 TTGGATGGACTTTTGGCTCAATATGGACAGTGGAAGTGGAACAGTGGACAAGTCAACACAGAC 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy  2041 CGAGGAGGGGCGGGGAAGTCAGCCAGGTTGCCGACGACCACCGAGACCCGCTCCGCC 2100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041 CGAGGAGGGGCGGGGAAGTCAGCCAGGTTGCCGACGACCACCGAGACCCGCTCCGCC 2100

Qy  2101 CCCAGGGGCTTCTGCGAGGCTTCAGCCATCCACTCCACTCGGATCTCTCCTCTGAA 2160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101 CCCAGGGGCTTCTGCGAGGCTTCAGCCATCCACTCCACTCGGATCTCTCCTCTGAA 2160

Qy  2161 CTCCCACGACGCTATCCCTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGC 2220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161 CTCCCACGACGCTATCCCTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGC 2220

Qy  2221 AAAATGCACACACCCTTTTCTGTGGCAAATCGTCTCGTACATGTGTACATATTAGAAA 2280
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221 AAAATGCACACACCCTTTTCTGTGGCAAATCGTCTCGTACATGTGTACATATTAGAAA 2280

Qy  2281 GGGAAGATGTTAAGATATGTGGCCCTGTGGGTTACACAGGGTGCCTGCAGCGCGGTAATATAT 2340
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281 GGGAAGATGTTAAGATATGTGGCCCTGTGGGTTACACAGGGTGCCTGCAGCGCGGTAATATAT 2340

Qy  2341 TTTAGAAATATATATCAATAACTAACTAACTCCAATTTTTAATCAATTATTAATTTT 2400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341 TTTAGAAATATATATCAATAACTAACTAACTCCAATTTTTAATCAATTATTAATTTT 2400

Qy  2401 TTTTTCTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG 2460
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401 TTTTTCTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG 2460

Qy  2461 GTCTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGAGGGAAAT 2520
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461 GTCTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGAGGGAAAT 2520

Qy  2521 CTCGTCGGAAGGACACTCACGGCAGTCTCGGATCACCTGTGTATGTCAACAGAGGGATA 2580
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521 CTCGTCGGAAGGACACTCACGGCAGTCTCGGATCACCTGTGTATGTCAACAGAGGGATA 2580

Qy  2581 CCGTCTCCTTGAAGAGGAAAACTCTGTCACTCGTCAGCTCGTCGTCGTCATACACCCATT 2640
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581 CCGTCTCCTTGAAGAGGAAAACTCTGTCACTCGTCAGCTCGTCGTCGTCATACACCCATT 2640

Qy  2641 TCTCTTTGCTTCACAGGTTTTAAACTGGTTTTGTTTTTGCATACTGCTATATAATTCTCTGTCT 2700
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641 TCTCTTTGCTTCACAGGTTTTAAACTGGTTTTGTTTTTGCATACTGCTATATAATTCTCTGTCT 2700

Qy  2701 CTCTCGTGTTTATTCTCTCCCCCTCCCCTTCTTCCTTCTTCTCCATTCCATTCTTTGAA 2760
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701 CTCTCGTGTTTATTCTCTCCCCCTCCCCTTCTTCCTTCTTCTCCATTCCATTCTTTGAA 2760

Qy  2761 TTTCCTCATCCCTCCATCTCAATCCCGTATCTACGGCACCCCCCCCCCAGGCAAAGCA 2820
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761 TTTCCTCATCCCTCCATCTCAATCCCGTATCTACGGCACCCCCCCCCCAGGCAAAGCA 2820

Qy  2821 GTGCTCTGAGTATCACATCACACACAAAGGAACAAAAGCGAAACACACAACCAGCTCAA 2880
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821 GTGCTCTGAGTATCACATCACACACAAAGGAACAAAAGCGAAACACACAACCAGCTCAA 2880

Qy  2881 CTTACACTTGGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAG 2940
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881 CTTACACTTGGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAG 2940

Qy  2941 AGGAAAACAGGAACCCACCACCAACCAACCAATCAACCAAACAAAGAAAAAATTCCACAATG 3000
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2941 AGGAAAACAGGAACCCACCACCAACCAACCAATCAACCAAACAAAGAAAAAATTCCACAATG 3000

Qy  3001 AAAGAATGTATTTGTCTTTGTCTTTTGCATTTGGTATAAGCCATCAATATTCAGCAAAATG 3060
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3001 AAAGAATGTATTTGTCTTTTGTCTTTTGCATTTGGTATAAGCCATCAATATTCAGCAAAATG 3060

Qy  3061 ATTCCTTTCTTTAAAAAAATGTGAGAAAGTAGAAAATTTACCAAGGTTGTTGGCC 3120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3061 ATTCCTTTCTTTAAAAAAATGTGAGAAAGTAGAAAATTTACCAAGGTTGTTGGCC 3120
```

```
Db   361 TTGGATGGACTTTGGCTCAATATGGACAGTGAGAATGTGGAACAAGTCAACACAGAC 420

Qy   421 ACAGAAACCGCCGTTGTCAAGGTCACATATGCAAGAGAAGAAGCAAAAATAGCCATG 480
         ||||||||||||||||||||||||||||| | | ||||||||||||||||||||||||
Db   421 ACAGAAACCGCCGTGTCAACGTCACATATGCACAGAGAAGAAGCAAAAATAGCCATG 480
     1501 ATCACCGGGCCACCCGGAAGCCCAGTTCAAGGACGAGGACGGATCTTTGGGAAACTGAAA 1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     1501 ATCACCGGGCCACCCGGAAGCCCAGTTCAAGGACGAGGACGGATCTTTGGGAAACTGAAA 1560

Qy   481 GAGAAGTCAAGGGGCATCAGTTTGAGAACTACTCCTTCTTCTTACATCCCGGAT 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GAGAAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCTTCTTACATCCCGGAT 540
     1561 GAGGAAACTTCTTAACCCAAAGAGAGAAGGTGGAAGCGCATATCAGAGTGCCC 1620
          ||||||||||||||||||||||||| |||||||||||||||||||||||||||
     1561 GAGGAAACTTCTTAACCCAAAGAAGAAGGTGGAAGCGCATATCAGAGTGCCC 1620

Qy   541 GAAGAGGTGAGCTCCCCTTCGCCCCCTCAGCGAGCCCAGCGTGGGACCACTCTTCCCG 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GAAGAGGTGAGCTCCCCTTCGCCCCCTCAGCGAGCCCAGCGTGGGACCACTCTTCCCG 600
     1621 TCTTCCACAGCTGCCGGGTGATTGGCAAAGGTGGCAAGACCGTGACGAACTGCGAAAC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     1621 TCTTCCACAGCTGCCGGGTGATTGGCAAAGGTGGCAAGACCGTGACGAACTGCGAAAC 1680

Qy   601 GAGCAAGGCCACGCCCTGGGGCACTTCTCAGCCCAGACAGATTGATTCCCGCTGCGG 660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GAGCAAGGCCACGCCCTGGGGCACTTCTCAGCCCAGACAGATTGATTCCCGCTGCGG 660
     1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTG 1740
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTG 1740

Qy   661 ATCCTGGTCCCACCCAGTTTGTTGGTGCCATCATCGGAAACTACTCCTTACATCCGGAT 720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 ATCCTGGTCCCACCCAGTTTGTTGGTGCCATCATCGGAAACTACTCCTTACATCCGGAT 720
     1741 ATCGTCAGAATTATCGGGCACTTCTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAA 1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     1741 ATCGTCAGAATTATCGGGCACTTCTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAA 1800

Qy   721 AACATCACTAAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAACTCTTGGACT 780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 AACATCACTAAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAACTCTTGGACT 780
     1801 ATTGTACAACAGTGAAGCAGCAGCAGAGAAATACCCTCAGGGAGTCGCCTCACAGCGC 1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     1801 ATTGTACAACAGTGAAGCAGCAGCAGAGAAATACCCTCAGGGAGTCGCCTCACAGCGC 1860

Qy   781 GCAGAGAAGCCTGTCACCATCATGCCACCCAGAGGGACTTCTGAAGCATGCGCATG 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GCAGAGAAGCCTGTCACCATCCATGCCACCCAGAGGGGACTTCTGAAGCATGCCGCATG 840
     1861 AGCAAGTGAGGCTCCCACAGCGACCACCAGGATGAATGTAGCCCTTCCAACAC 1920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
     1861 AGCAAGTGAGGCTCCCACAGCGACCACCAGGATGAATGTAGCCCTTCCAACAC 1920

Qy   841 ATTCTTGAAATCATGCAGATGAGACCAAATCAGCTAGCAGGAGATTCCTCTG 900
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 ATTCTTGAAATCATGCAGATGAGACCAAATCAGCTAGCAGGAGATTCCTCTG 900
     1921 CTGACAGAATGAGACCAAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAG 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     1921 CTGACAGAATGAGACCAAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAG 1980

Qy   901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAGAAGGCAGAAATTTG 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAGAAGGCAGAAATTTG 960
     1981 GAATGAGAGTCTGCGGAGGCGGCCCAGGAGACTCTGCGGAGGCGGCCCAGGAGAACCCCAGGGGC 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     1981 GAATGAGAGTCTGCGGAGGCGGCCCAGGAGACTCTGCGGAGGCGGCCCAGGAGAACCCCAGGGGC 2040

Qy   961 AAGAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 AAGAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020
     2041 CGAGGAGGGGCCGGGAGGGGCCGGGAAGGTCAGCCAGGTTGCCAGAACCACCGAGCCCTCCCGCC 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     2041 CGAGGAGGGGCCGGGAGGGGCCGGGAAGGTCAGCCAGGTTGCCAGAACCACCGAGCCCTCCCGCC 2100

Qy  1021 ATATACAACCCGGAAAGAACCATCCACACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCT 1080
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 ATATACAACCCGGAAAGAACCATCCACACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCT 1080
     2101 CCCCAGGGCTTCTGCAGGCTTCAGCTGAGCCTCCACTTCACCACTCGGATCTCTCCTGAA 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     2101 CCCCAGGGCTTCTGCAGGCTTCAGCTGAGCCTCCACTTCACCACTCGGATCTCTCCTGAA 2160

Qy  1081 GAGATAGAGATTATGAAGAGCTGCGTGAGGAGGCCTTGAAAATATGTCGGCTGTTAAC 1140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 GAGATAGAGATTATGAAGAGCTGCGTGAGGAGGCCTTGAAAATATGTCGGCTGTTAAC 1140
     2161 CTCCCACGACGCTATCCTTTTAGTTGAACTAACATTAGGTGAACGTGTTCAAAGCCAAGC 2220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     2161 CTCCCACGACGCTATCCTTTTAGTTGAACTAACATTAGGTGAACGTGTTCAAAGCCAAGC 2220

Qy  1141 CAACAAGGCCAATCTGATCCCAGGTTGAACCTTCAGCCGCACTTGCATCTTTTCAACAGGA 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 CAACAAGGCCAATCTGATCCCAGGTTGAACCTTCAGCCGCACTTGCATCTTTTCAACAGGA 1200
     2221 AAAATGCACACCCCTTTTCTGTGGCAAATCGTCTGTGTACATGTGTTGTACATATTAGAAA 2280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     2221 AAAATGCACACCCCTTTTTCTGTGGCAAATGTCTGTGTACATGTGTTGTACATATTAGAAA 2280

Qy  1201 CTGTCCGTGCTATCTCCACCAGCAGGGGCCCCCGCTGCCCCCTACCAC 1260
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 CTGTCCGTGCTATCTCCACCAGCAGGGGCCCCCGCTGCCCCCTACCAC 1260
     2281 GGGAAGATGTTAAGATATGTGCCTGTGGGTTACACAGGGTGCCTGCCTGTGCAGCCGTAATATAT 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     2281 GGGAAGATGTTAAGATATGTGCCTGTGGGTTACACAGGGTGCCTGCCTGTGCAGCCGTAATATAT 2340

Qy  1261 CCCTTCACTACACTCCGATACTTCTTCAGCCTGTACCCCGTTGGGCCCG 1320
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 CCCTTCACTACACTCCGATACTTCTTCAGCCTGTACCCCGTTGGGCCCG 1320
     2341 TTTAGAAATAATATCAAATAACTCAACTCCAATTTTAATCAATTATTAATTTT 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||| |||
     2341 TTTAGAAATAATATCAAATAACTCAACTCCAATTTTTAATCAATTATTAATTTT 2400

Qy  1321 TTCCCGCATCATCACTCCACCACTCCAGAGCAGGAGATTGGAATCTCTTCATCCCAACCCAG 1380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 TTCCCGCATCATCACTCCACCACTCCAGAGCAGGAGAGTTGGAATTCTTCATCCCAACCCAG 1380
     2401 TTTTTCTTTTTAAAGAGAAAGCAGGCTTTTCTCGAGACTTTAAAGAGATTAAAGTCTTTGGGAG 2460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     2401 TTTTTCTTTTTAAAGAGAAAGCAGGCTTTTCTCGAGACTTTAAAGAGAATTAAAGTCTTTGGGAG 2460

Qy  1381 GCTGTGGGGCGCATCGGGAAGAAAGGGGCACACATCAAACAGCTGGCAGAGATTCGCC 1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 GCTGTGGGGCGCATCGGGAAGAAAGGGGCACACATCAAACAGCTGGCAGAGATTCGCC 1440
     2461 GTCTCACGGTGTAGAGAGAGGGCACCCGCCACCCGGAGATTCGCC 2520
          |||||||||||||||||||||||||||||||||||||||||||||
     2461 GTCTCACGGTGTAGAGAGAGGGACTTTGAGGGCCACCAACCCGGAGATTCGCC 2520

Qy  1441 GGAGCCCTTATCAAGATTGCCCCCTGCCCGGAATTGCCCCCTGTCATCCGGTCAATGTCATC 1500
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 GGAGCCCTTATCAAGATTGCCCCCTGCCCGGAATTGCCCCCTGTCATCCGGTCAATGGTCATC 1500
     2521 CTCGTCGGAAGGACACTCACGGCAGTTCTTGGATCGTGTATGTCAACAGAAGGGATA 2580
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     2521 CTCGTCGGAAGGACACTCACGGCAGTTCTTGGATCGTGTATGTCAACAGAAGGGATA 2580
```

JOURNAL   antigens, these antigens and method of using the same
           Patent: JP 2002512049-A 4 23-APR-2002;
           LUDWIG INSTITUTE FOR CANCER RESEARCH
COMMENT   OS   Homo sapiens (human)
           PN   JP 2002512049-A/4
           PD   23-APR-2002
           PF   16-MAR-1999 JP 2000545030
           PR   17-APR-1998 US   09/061709
           PI   YAO TSENG CHEN,ALI GURE,SOLAM TSANG,ELISABETH STOCKERT,ELKE
           PI   JAGER,
           PI   ALEXANDER KNUTH,LLOYD J OLD
           PC   C12N15/09,A61K35/12,A61K39/00,A61P35/39,A61P35/00,C07K16/32,
           PC   C12N1/15,
           PC   C12N1/19,C12N1/21,C12N5/10,C12P21/08,C12Q1/68,G01N33/53,G01N33/ PC
           574,
           PC   C12N15/00,C12N5/00
           CC   n=not determined(3372)
           FH   Key   Location/Qualifiers
           FT   source   1..3412
           FT   /organism='Homo sapiens (human)'.
FEATURES   Location/Qualifiers
    source   1..3412
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
ORIGIN

Query Match   100.0%;   Score 3411;   DB 6;   Length 3412;
Best Local Similarity   100.0%;   Pred. No. 0;
Matches 3412;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   1 GGCAGCGGAGGAGGCGAGAGCGCCGGGTACCCGGCCGGGGGAGCCGGGCCTCGGGG   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGCAGCGGAGGAGGCGAGAGCGCCGGGTACCCGGCCGGGGGAGCCGGGCCTCGGGG   60

Qy   61 AAGAGACGGATGATGAACAAGCTTTACATCGGGACAGAAGCTGCCCGACGCCGCCGAC   120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AAGAGACGGATGATGAACAAGCTTTACATCGGGACAGAAGCTGCCCGACGCCGCCGAC   120

Qy   121 GACCTCCGGCAGCTCTTTGGGGACCAGGAAGCTGCCCGGCGGGACAGTGCCTGCTGAAG   180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GACCTCCGGCAGCTCTTTGGGGACCAGGAAGCTGCCCGGCGGGACAGTGCCTGCTGAAG   180

Qy   181 TCCGGCTACGCCTTCGTGGACTACCCCGACCACCAGAACTGGGCCATCCGGCCATCGAGACC   240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TCCGGCTACGCCTTCGTGGACTACCCCGACCACCAGAACTGGGCCATCCGGCCATCGAGACC   240

Qy   241 CTCTCGGGTAAAGTGGAATTGCATGGGAAATCAGATTGGAAATCAGTTGATTACTCAGTCTCTAA   300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CTCTCGGGTAAAGTGGAATTGCATGGGAAATCAGATTGGAAATCAGTTGATTACTCAGTCTCTAA   300

Qy   301 AAGTAAGGACGAGAAATTCAGATTGAAACATCCCCTCCACCTGCCAGTGGGAGGTG   360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AAGTAAGGACGAGAAATTCAGATTGAAACATCCCCTCCACCTGCCAGTGGGAGGTG   360

Qy   361 TTGGATGGACTTTTGGCTCAATATGGGACAGTGGAACAAGTGGAACAAGTCAACACAGAC   420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TTGGATGGACTTTTGGCTCAATATGGGACAGTGGAACAAGTGGAACAAGTCAACACAGAC   420

Qy   421 ACAGAAACCGCCGTTGTCAACGTCACATATGCACATATGCACATATAGCCATG   480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ACAGAAACCGCCGTTGTCAACGTCACATATGCACATATGCACATATAGCCATG   480

Qy   481 GAGAAGCTAAGGCGGGCATCAGTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGAT   540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GAGAAGCTAAGGCGGGCATCAGTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGAT   540

Qy   541 GAAGAGGTGAGCTCCCCTTCGCCCCTCCAGCGTGGGGACCACTCTTCCCGG   600
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GAAGAGGTGAGCTCCCCTTCGCCCCTCCAGCGTGGGGACCACTCTTCCCCG   600

Qy   601 GAGCAAGGCCACGCCGCCCTGGGGCCACTTCTCAGGCCAGCAGATTGATTCCCGCTGCGG   660

Qy   2581 CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTCCTCATGCCTGTCTCAGCTCATACACCCATT   2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTCCTCATGCCTGTCTCAGCTCATACACCCATT   2640

Qy   2641 TCTCTTGCTTCACAGGTTTAAACTGGTTTTTGCATATCGTCATATATAATTCTCTGTCT   2700
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 TCTCTTGCTTCACAGGTTTAAACTGGTTTTTGCATATCGTCATATATAATTCTCTGTCT   2700

Qy   2701 CTCTCTGTTTATCTCCTCCCCTCCCCCCCCCCCCCCCTTCTTCCATCTCCATTCCATTCTTTGAA   2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 CTCTCTGTTTATCTCCTCCCCTCCCCCCCCCCCCCCCTTCTTCCATCTCCATTCCATTCTTTGAA   2760

Qy   2761 TTTCCTCATCCCTCCATCTCAATCCGTATCTACGCACCCCCCCCCCCCCCCCCAGGCAAAGCA   2820
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 TTTCCTCATCCCTCCATCTCAATCCGTATCTACGCACCCCCCCCCCCCCCCCCAGGCAAAGCA   2820

Qy   2821 GTGCTCTGAGTATCATCATCACACAAAAGGAACAAAAGCGAAACAAACCAGCCTCAA   2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 GTGCTCTGAGTATCATCATCACACAAAAGGAACAAAAGCGAAACAAACCAGCCTCAA   2880

Qy   2881 CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAG   2940
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAG   2940

Qy   2941 AGGAAAACAGGAAACCACCAACCAAAAACTCAACCAAATCAACCAAACAAAGAAAAAATTCCACAATG   3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 AGGAAAACAGGAAACCACCAACCAAAAACTCAACCAAATCAACCAAACAAAGAAAAAATTCCACAATG   3000

Qy   3001 AAAGAATGTATTTTGTCTTTTTGCATTTGTGTATAAGCCATCAATATTCAGCAAAATG   3060
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 AAAGAATGTATTTTGTCTTTTTGCATTTGTGTATAAGCCATCAATATTCAGCAAAATG   3060

Qy   3061 ATTCCTTTCTTTAAAAAAAAATGTGGAGGAAAGTAGAAAATTTACCAAGGTTGTTGGCC   3120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 ATTCCTTTCTTTAAAAAAAAATGTGGAGGAAAGTAGAAAATTTACCAAGGTTGTTGGCC   3120

Qy   3121 CAGGGCGTTAAATTCACAGATTTTTTTAACGAGAAAACACAGAAGAAGCTACCTCAG   3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 CAGGGCGTTAAATTCACAGATTTTTTTAACGAGAAAACACAGAAGAAGCTACCTCAG   3180

Qy   3181 GTGTTTTACCTGAAGTGACACCTTGCTCTGTGTTCCCTTAGAGATTTGTAAAGTGATAG   3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 GTGTTTTACCTGAAGTGACACCTTGCTCTGTGTTCCCTTAGAGATTTGTAAAGTGATAG   3240

Qy   3241 TTGGAGCATTTTTATTTTTTTAATAAAATGAGTTGGAAAAATAAGAGTTGGAAAAATAAGAATATCAACT   3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 TTGGAGCATTTTTATTTTTTTAATAAAATGAGTTGGAAAAATAAGAGTTGGAAAAATAAGAATATCAACT   3300

Qy   3301 GCCAGCCTGGAGAAGTGACAGTCCAAGTGTGCAACAGCTGTTCTCGAATTGTCTTCCGCT   3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 GCCAGCCTGGAGAAGTGACAGTCCAAGTGTGCAACAGCTGTTCTCGAATTGTCTTCCGCT   3360

Qy   3361 AGCCAAGAACCNATATGGCCTTCTTTTTGGACAAACCTTGAAAATGTTTATTT   3412
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 AGCCAAGAACCNATATGGCCTTCTTTTTGGACAAACCTTGAAAATGTTTATTT   3412

RESULT 3
BD209925
LOCUS   BD209925   3412 bp   DNA   linear   PAT 17-JUL-2003
DEFINITION Isolated nucleic acid molecules encoding cancer-associated
            antigens, these antigens and method of using the same.
ACCESSION   BD209925
VERSION   BD209925.1   GI:33019695
KEYWORDS   JP 2002512049-A/4.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3412)
  AUTHORS   Chen,Y.T., Gure,A., Tsang,S., Stockert,E., Jager,E., Knuth,A. and
            Old,L.J.
  TITLE   Isolated nucleic acid molecules encoding cancer-associated

```
Db   1681 TTAACCAGTGCAGAAGTCATCGTGCCTTCGTGACCAAACGCCAGATGAAAATGAGGAAGTG 1740
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1741 ATCGTCAGAATTATCGGGCACTTCTTGTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAA 1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 ATCGTCAGAATTATCGGGCACTTCTTGTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAA 1800

Qy   1801 ATTGTACAACAGGTGAGCGAGCGAGCAGGAGAAATACCCTCAGGAGGTCGCCTCACAGCGC 1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 ATTGTACAACAGGTGAGCGAGCGAGCAGGAGAAATACCCTCAGGAGGTCGCCTCACAGCGC 1860

Qy   1861 AGCAAGTGAGGCTCCCACAGGCACCAGCAAAACAACGATGAATGTAGCCTTCCAACAC 1920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 AGCAAGTGAGGCTCCCACAGGCACCAGCAAAACAACGATGAATGTAGCCTTCCAACAC 1920

Qy   1921 CTGACAGAATGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAG 1980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CTGACAGAATGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAG 1980

Qy   1981 GAATGAGAAGTCTGCGGAGGCGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGC 2040
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GAATGAGAAGTCTGCGGAGGCGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGC 2040

Qy   2041 CGAGGAGGGCGGGAAGGTCAGCGAGGTTTGCCAGAACCACCGAGCCCCGCCTCCCGCC 2100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 CGAGGAGGGCGGGAAGGTCAGCGAGGTTTGCCAGAACCACCGAGCCCCGCCTCCCGCC 2100

Qy   2101 CCCCAGGGCTTCTGCGAGGCTTCAGCGCCATCCACTCACCACTCGGATCCTCCTGAA 2160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 CCCCAGGGCTTCTGCGAGGCTTCAGCGCCATCCACTCACCACTCGGATCCTCCTGAA 2160

Qy   2161 CTCCCACGACGCTATCCTTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGC 2220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CTCCCACGACGCTATCCTTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGC 2220

Qy   2221 AAAATGCACACCCTTTTCTGTGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAA 2280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 AAAATGCACACCCTTTTCTGTGCAAATCGTCTCTGTACATGTGTACATATTAGAAA 2280

Qy   2281 GGGAAGATGTTAAGATATGTGGCCTGTGGGTACACAGGGGTGCCTGCAGCGGTAATATAT 2340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 GGGAAGATGTTAAGATATGTGGCCTGTGGGTACACAGGGGTGCCTGCAGCGGTAATATAT 2340

Qy   2341 TTTAGAAATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTT 2400
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 TTTAGAAATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTT 2400

Qy   2401 TTTTTCTTTTTAAAGAAGAAAGCAGGCTTTTGAGGCCACCCGACAAAATTCACCCAGAGGGAAAT 2460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 TTTTTCTTTTTAAAGAAGAAAGCAGGCTTTTGAGGCCACCCGACAAAATTCACCCAGAGGGAAAT 2460

Qy   2461 GTCTCACGGTGTAGAGAGGACACTCACGGCAGTTCTGATCACCTGTGTATGTCAACAGGGATA 2520
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 GTCTCACGGTGTAGAGAGGACACTCACGGCAGTTCTGATCACCTGTGTATGTCAACAGGGATA 2520

Qy   2521 CCGTCTCCTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCT 2580
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 CCGTCTCCTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCT 2580

Qy   2581 CCGTCTCCTTGAAGAGGAAACTCTGTCACTCTGCCTGTCAGTTGTGATCGTCAGTTGTGATCAAACCACCATT 2640
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 CCGTCTCCTTGAAGAGGAAACTCTGTCACTCTGCCTGTCAGTTGTGATCGTCAGTTGTGATCAAACCACCATT 2640

Qy   2641 CTCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGGTATATATATTCTCTGTCT 2700
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 CTCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGGTATATATATTCTCTGTCT 2700

Qy   2701 CTCTCTGTTTATCTCTCCCCTTCTCTCCCATCTCCATCTCCATTCTTTGAA 2760
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 CTCTCTGTTTATCTCTCCCCTTCTCTCCCATCTCCATCTCCATTCTTTGAA 2760

Qy   2761 TTTCCTCATCCCTCCATCTCAATCCCTCAATCCCGTATCCCGTATCTCAATCCGCAAAGCA 2820
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 TTTCCTCATCCCTCCATCTCAATCCCTCAATCCCGTATCCCGTATCTCAATCCGCAAAGCA 2820
```

```
Db    601 GAGCAAGGCCACGCCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCCG 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    661 ATCCTGGTCCCCACCCAGTTTGTTGTCGTGCCATCATCGGAAGGAGGGCTTGACCATAAAG 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ATCCTGGTCCCCACCCAGTTTGTTGTCGTGCCATCATCGGAAGGAGGGCTTGACCATAAAG 720

Qy    721 AACATCACTAAGCAGCACCCAGTCCCGGGTAGATATCCATAGAAAAAGAGAACTCTGGAGCT 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AACATCACTAAGCAGCACCCAGTCCCGGGTAGATATCCATAGAAAAAGAGAACTCTGGAGCT 780

Qy    781 GCAGAGAAGCCTGTCACCATCCATGCCACCCAGAGGGGACTTTGAAGCATGCCGCATG 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GCAGAGAAGCCTGTCACCATCCATGCCACCCAGAGGGGACTTTGAAGCATGCCGCATG 840

Qy    841 ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG 900

Qy    901 AAAATCTTGGCACACAAATGGCTTGTTGGTTGGAAGACTGATTGGAAAAAGAGGCAGAAATTTG 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 AAAATCTTGGCACACAAATGGCTTGTTGGTTGGAAGACTGATTGGAAAAAGAGGCAGAAATTTG 960

Qy    961 AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020

Qy   1021 ATATACAACCCGGAAAGAACCATCACTGTTGAGGGCACAGTCCTGTGCCAGTGCT 1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ATATACAACCCGGAAAGAACCATCACTGTTGAGGGCACAGTCCTGTGCCAGTGCT 1080

Qy   1081 GAGATAGAGATTATGAAGAAGCTCGCTGAGAAGCCTTTGAAAATATGTCTGGCTGTTAAC 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GAGATAGAGATTATGAAGAAGCTCGCTGAGAAGCCTTTGAAAATATGTCTGGCTGTTAAC 1140

Qy   1141 CAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGA 1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGA 1200

Qy   1201 CTGTCCGTGTCTATCTCCACCACAGAGGCCCCGGGAGCTCCCCGGCCCCCGCGTACCAC 1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CTGTCCGTGTCTATCTCCACCACAGAGGCCCCGGGAGCTCCCCGGCCCCCGCGTACCAC 1260

Qy   1261 CCCTTCACTCACTACCCACTCGGGATACTTCTCCAGCTCGTACCCCGTGTACCCATCACCAGTTGGGCCCG 1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CCCTTCACTCACTACCCACTCGGGATACTTCTCCAGCTCGTACCCCGTGTACCCATCACCAGTTGGGCCCG 1320

Qy   1321 TTCCCGCATCATCACTCTTATCCGAGCAGGAGCAGGAGAGATTGTGAATCTTCTTCATCCCCAACCCAG 1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TTCCCGCATCATCACTCTTATCCGAGCAGGAGCAGGAGAGATTGTGAATCTTCTTCATCCCCAACCCAG 1380

Qy   1381 GCTGTGGGGCCATCATCGGGAGAAGAGGGGCACACATCAAACAGCTTGGCGAGATTGCC 1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GCTGTGGGGCCATCATCGGGAGAAGAGGGGCACACATCAAACAGCTTGGCGAGATTGCCC 1440

Qy   1441 GGAGGCCTCTATCAAGATTGCCCCTGCGGAAGCCCAGAGCGTCAGCGAAGGATGGTCATC 1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GGAGGCCTCTATCAAGATTGCCCCTGCGGAAGCCCAGAGCGTCAGCGAAGGATGGTCATC 1500

Qy   1501 ATCACCGGGGCCACCGGAAGCCCAGTTCAAGGCCCCGGACGGATCTTTGGGAAACTGAAA 1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 ATCACCGGGGCCACCGGAAGCCCAGTTCAAGGCCCCGGACGGATCTTTGGGAAACTGAAA 1560

Qy   1561 GAGGAAAACTTCTTTAACCCCAAAGAAGAGTGAAGCTGGAAGCGCCATATCAGAGTGCCC 1620
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 GAGGAAAACTTCTTTAACCCCAAAGAAGAGTGAAGCTGGAAGCGCCATATCAGAGTGCCC 1620

Qy   1621 TCTTCCACAGCTGGCCGGGTGATTGGCAAAGTGTGCAAGACCGTGACAACTGCAGAAC 1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TCTTCCACAGCTGGCCGGGTGATTGGCAAAGTGTGCAAGACCGTGACAACTGCAGAAC 1680

Qy   1681 TTAACCAGTGCAGAAGTCATCGTGCCTTCGTGACCAAACGCCAGATGAAAATGAGGAAGTG 1740
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy   2821  GTGCTCTGAGTATCACATCACACAAAGGAACACAAAGGAACACAAAACCAGCCTCAA  2880
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821  GTGCTCTGAGTATCACATCACACAAAGGAACACAAAGGAACACAAAACCAGCCTCAA  2880

Qy   2881  CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAG  2940
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881  CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAG  2940

Qy   2941  AGGAAAACAGGAACCCACCACCAACCAATCACCCAACAAAGAAAAATTCCACAATG  3000
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941  AGGAAAACAGGAACCCACCACCAACCAACCAATCACCCAACAAAGAAAAATTCCACAATG  3000

Qy   3001  AAAGAATGTATTTTGTCTTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATG  3060
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001  AAAGAATGTATTTTGTCTTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATG  3060

Qy   3061  ATTCCTTCTTTAAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAGGTTGTTGGCC  3120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061  ATTCCTTCTTTAAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAGGTTGTTGGCC  3120

Qy   3121  CAGGGCGTTAAATTCACAGATTTTTTTTTAACGAGAAAAACACAGAAGAAGCTACCTCAG  3180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121  CAGGGCGTTAAATTCACAGATTTTTTTTTTAACGAGAAAAACACAGAAGAAGCTACCTCAG  3180

Qy   3181  GTGTTTTTACCTCAGCACCTTGCTCTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAG  3240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181  GTGTTTTTACCTCAGCACCTTGCTCTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAG  3240

Qy   3241  TTGGAGCATTTTTTTTTATTTTTTATAAAAATGAGTGGAAAAATAAGATATCAACT  3300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241  TTGGAGCATTTTTTTTTTATTTTTTTATAAAAATGAGTGGAAAAAAATAAGATATCAACT  3300

Qy   3301  GCCAGCCTGGAGAAGTGACAGTCCAAGTGCGCAACAGTGTTGTTCGAATTGTCTTCCGCT  3360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301  GCCAGCCTGGAGAAGTGACAGTCCAAGTGCGCAACAGTGTTGTTCGAATTGTCTTCCGCT  3360

Qy   3361  AGCCAAGAACCNATATGGCCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT  3412
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361  AGCCAAGAACCNATATGGCCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT  3412
```

RESULT 4
BC021290
LOCUS   BC021290   3633 bp   mRNA   linear   PRI 04-OCT-2003
DEFINITION Homo sapiens IGF-II mRNA-binding protein 2, mRNA (cDNA clone
            MGC:29539 IMAGE:5090334), complete cds.
ACCESSION   BC021290
VERSION   BC021290.2 GI:33870841
KEYWORDS   MGC.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3633)
AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE   22388257
PUBMED   12477932
REFERENCE   2 (bases 1 to 3633)
AUTHORS   Strausberg,R.
TITLE   Direct Submission
JOURNAL   Submitted (14-JAN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   On Aug 19, 2003 this sequence version replaced gi:18204200.
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Matrian,S.D., McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 39 Row: o Column: 21
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 27552765.
FEATURES   Location/Qualifiers
     source   1..3633
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:29539 IMAGE:5090334"
                     /tissue_type="Pancreas; epithelioid carcinoma"
                     /clone_lib="NIH_MGC_42"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     gene   1..3633
                     /gene="IMP-2"
                     /note="synonym: VICKZ2"
                     /db_xref="LocusID:10644"
     CDS   68..1864
                     /codon_start=1
                     /product="IGF-II mRNA-binding protein 2"
                     /protein_id="AAH21290.1"
                     /db_xref="GI:18204201"
                     /db_xref="LocusID:10644"
                     /translation="MNKLYIGNLSPAVTADDLRQLFGDRKLPLAGQVLLKSGYAFVDY
                     PDQNWAIRAIETLSGKVELHGKIMEVDYSVSKKLRSRKIQRNIPPHLQWEVLDGLLA
                     QYGTVENVEQVNTDTETAVVNVYATREEAKIAMEKLSGHQFENYSFKISYIPDEEVS
                     SPSPPQRAQRGDHSSREQGHAPGGTSQARQIDFPLRILVPTQFVGAIIGKEGLTIKNI
                     TKQTGSRVDIHRKENSGAAEKPVTIHATPEGTSEACRMILEIMQKEADETKLAEEIPL
                     KILAHNGLVGRLIGKEGRNLKKIEHETGKTITISSLQDLSIYNPERTITVKGTVEACA
                     SAEIEIMKKLREAFENDMLAVNQQANLIPGLNLSALGIFSTGLSVLSPPAGPRGAPPA
                     APYHPFTHSGYFSSLYPHQFGPFHHSYPEQEIVNLFIPQAVGAIIGKKGAHIK
                     QLARFAGASIKIAPAEGPDVSERMVIITGPPEAQFKAQGRIFGKLKEENFFNPKEEVK
                     LEAHIRVPSSTAGRVIGKGGKTVNELQNLTSAEVIVPRDQTPDENEEVIVRIIGHFFA
                     SQTAQRKIREIVQQVKQQEQKYPQGVASQRSK"
     misc_feature   74..280
                     /note="RRM; Region: RNA recognition motif"
                     /db_xref="CDD:smart00360"
     misc_feature   650..853
```

us-09-270-437d-6.rge

```
misc_feature        /note="KH; Region: K homology RNA-binding domain"
                    /db_xref="CDD:smart00322"
                    1340..1555
                    /note="KH; Region: K homology RNA-binding domain"
                    /db_xref="CDD:smart00322"
ORIGIN

Query Match         99.2%;  Score 3383;  DB 9;  Length 3633;
Best Local Similarity   99.9%;  Pred. No. 0;
Matches 3405;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    2;
```

```
Qy        6  CGGGAGGAGGCGAGGAGGCGCGGGTACCCGGCCGGGGGAGCCGCGGGCTCTCGGGGAAGAG   65
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1  CGGGAGGAGGCGAGGAGGCGCGGGTACCCGGCCGGGGGAGCCGCGGGCTCTCGGGGAAGAG   60

Qy       66  ACGGATGATGAACAAGCTTTACATCGGGAACTGCCCCGTCACCGCCGACGACCT          125
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61  ACGGATGATGAACAAGCTTTACATCGGGAACTGCCCCGTCACCGCCGACGACCT          120

Qy      126  CCGGCAGCTCTTTGGGGACAGGAAGTGCCCCTGCTGGGACAGGTCCTGCTGAAGTCCGG    185
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121  CCGGCAGCTCTTTGGGGACAGGAAGTGCCCCTGCTGGGACAGGTCCTGCTGAAGTCCGG    180

Qy      186  CTACGGCCTTCGTGGACTACCCCGACCAGGAACTGGGCCATCCGCGCCAGACCCCTCTC    245
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181  CTACGGCCTTCGTGGACTACCCCGACCAGGAACTGGGCCATCCGCGCCAGACCCCTCTC    240

Qy      246  GGGTAAAGTGGAATTGCATGGGAAAATCATGGGAAGTTGATTACTCAGTTCTAAAAAGCT    305
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241  GGGTAAAGTGGAATTGCATGGGAAAATCATGGGAAGTTGATTACTCAGTTCTAAAAAGCT    300

Qy      306  AAGGAGCAGGAAAATTCAGATTCGAAACATCCCTCCTCCACCTGGCAGTGGAGTGTTGGA    365
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301  AAGGAGCAGGAAAATTCAGATTCGAAACATCCCTCCTCCACCTGGCAGTGGAGTGTTGGA    360

Qy      366  TGGACTTTTGGCTCAATATGGACAGTGGAGAATGTGGAACAAGTCAACACAGACACAGA    425
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361  TGGACTTTTGGCTCAATATGGACAGTGGAGAATGTGGAACAAGTCAACACAGACACAGA    420

Qy      426  AACCGCCGTTGTCAACGTCACATATGCACATATGGCATTCTTCAGGACCAATAGCCATGGAGAA   485
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421  AACCGCCGTTGTCAACGTCACATATGCACATATGGCATTCTTCAGGACCAATAGCCATGGAGAA   480

Qy      486  GCTAAGCGGGCATCAGTTTGAGAACTACTCCCTCACATCCCCGATGAAGA    545
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db      481  GCTAAGCGGGCATCAGTTTGAGAACTACTCCCTCACATCCCCGATGAAGA    540

Qy      546  GGTGAGCTCCCCTTCGCCCCTCAGCGAGCCCAGCGTGGGGACCACTCTTCCCGGGAGCA    605
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541  GGTGAGCTCCCCTTCGCCCCTCAGCGAGCCCAGCGTGGGGACCACTCTTCCCGGGAGCA    600

Qy      606  AGGGCCACCAGCGCCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTCCCGCTGCGGATCCT    665
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601  AGGGCCACCAGCGCCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTCCCGCTGCGGATCCT    660

Qy      666  GGTCCCACCAGTTGTTGGTGGTGCCATCATCGGAAGAGGGCTTGACCATAAAGAACAT    725
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661  GGTCCCACCAGTTGTTGGTGGTGCCATCATCGGAAGAGGGCTTGACCATAAAGAACAT    720

Qy      726  CACTAAGCAGACCCAGTCCGGGTAGATATCCCCAGAGGGGACTTCTGAAGCTGCGGAGA    785
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721  CACTAAGCAGACCCAGTCCGGGTAGATATCCCCAGAGGGGACTTCTGAAGCTGCGGAGA    780

Qy      786  GAAGCCTGTCACCATCCACCATGCCACCCCGAGGGGACTTCTGAAGCATGCCGCATGATTCT    845
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db      781  GAAGCCTGTCACCATCCACCATGCCACCCCGAGGGGACTTCTGAAGCATGCCGCATGATTCT    840

Qy      846  TGAAATCATGCAGAAAGAGGCAGATGACCAAACTAGCCAGACAGAATTCCTCTGAAAAT    905
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841  TGAAATCATGCAGAAAGAGGCAGATGACCAAACTAGCCAGACAGAATTCCTCTGAAAAT    900

Qy      906  CTTGGCCACAATGGCTTGGTTGGTTGAAGACTGATTGAGAAGAAGGCAGAAATTGAAGAA    965
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901  CTTGGCCACAATGGCTTGGTTGGTTGAAGACTGATTGAGAAGAAGGCAGAAATTGAAGAA    960
```

```
Qy      901  CTTGGCACAATGGCTTGGTTGGTTGAAGACTGATTGAGAAGAAGGCAGAAATTGAAGAA    960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      966  AATTGAACATGAAACAGGAGCAGGAGACCAAGATAACAATCTCATCTTTGCAGGATTGAGCATATA   1025

Db      961  AATTGAACATGAAACAGGAGCAGGAGACCAAGATAACAATCTCATCTTTGCAGGATTGAGCATATA   1020

Qy     1026  CAACCCGGAAAGAACCATCACTGTGAAGGACCACAGTTGAGGCCTGTGCCAGTGCTGAGAT   1085
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021  CAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGAGAT   1080

Qy     1086  AGAGATTATGAAGAAGTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTGTTAACCAACA   1145
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081  AGAGATTATGAAGAAGTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTGTTAACCAACA   1140

Qy     1146  AGCCAATCGATCCCAGGGTTGAACCTCAGCGCCACTTGGCATCTTTTCAACAGGACTGTC   1205
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141  AGCCAATCGATCCCAGGGTTGAACCTCAGCGCCACTTGGCATCTTTTCAACAGGACTGTC   1200

Qy     1206  CGTGCTATCTCCACCAGCAGGGCCCCGGGGAGCTCCCCCGTGCCCCTACCACCCCTT   1265
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201  CGTGCTATCTCCACCAGCAGGGCCCCGGGGAGCTCCCCCGTGCCCCCTACCACCCCTT   1260

Qy     1266  CACTACCCACTCCGATACTTCTCCAGCCTGTACCCGTGAGCCCATCACACCAGTTGGCCCGTTCCC   1325
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1261  CACTACCCACTCCGATACTTCTCCAGCCTGTACCCGTCCCCATCACCAGTTGGCCCGTTCCC   1320

Qy     1326  GCATCATCACTCCTATCGACGCAGCAGGAGATTGTGAATCTCTTCATCCCAACCCAGGCTGT   1385
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1321  GCATCATCACTCTTATCGACGCAGCAGGAGATTGTGAATCTCTTCATCCCAACCCAGGCTGT   1380

Qy     1386  GGGCGCCATCATCGGGAAGACAGAGAGGGGCACACATCAAACAGCTGGCGAGATTCGCCCGAGC   1445
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1381  GGGCGCCATCATCGGGAAGACAGAGAGGGGCACACATCAAACAGCTGGCGAGATTCGCCCGAGC   1440

Qy     1446  CTCTATCAAGATTGCCCCTGCCCGGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCATCAC   1505
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1441  CTCTATCAAGATTGCCCCTGCCCGGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCATCAC   1500

Qy     1506  CGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGACCGGATCTTTGGGAAACTGAAAGAGAA   1565
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1501  CGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGACCGGATCTTTGGGAAACTGAAAGAGAA   1560

Qy     1566  AAACTTCTTTAACCCAAAGAAGAACTGAGCTGCAAGGTGGAAGGCGCATATCAGAGTGCCCTCTTC   1625
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1561  AAACTTCTTTAACCCAAAGAAGAACTGAGCTGCAAGGTGGAAGGCGCATATCAGAGTGCCCTCTTC   1620

Qy     1626  CACAGCTGGCGGGGTGATTGGCAAGGTGGCAAGACCGTGAACGAACTGCAGAACTTAAC   1685
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1621  CACAGCTGGCGGGGTGATTGGCAAGGTGGCAAGACCGTGAACGAACTGCAGAACTTAAC   1680

Qy     1686  CAGTGCAGAAGTCATCGTGCCTGTGACCAAAATGAGGAGAGTGATCGT   1745
             ||||||||||||||||||||||||||||||||||||||||||||||
Db     1681  CAGTGCAGAAGTCATCGTGCCTGTGACCAAAATGAGGAGAGTGATCGT   1740

Qy     1746  CAGAATTATCGGGCACTTCTTTGCTAGCTAGCAGCAGACTGCACAGCGCACAGGAATTGT   1805
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1741  CAGAATTATCGGGCACTTCTTTGCTAGCTAGCAGCAGACTGCACAGCGCACAGGAATTGT   1800

Qy     1806  ACAACAGGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACGCCAGCAA   1865
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1801  ACAACAGGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACGCCAGCAA   1860

Qy     1866  GTGAGGCTCCCACAGCACCAGGAGCACCAGCAGCCTGAATGTAGCCCTTCCAACACCTGAC   1925
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1861  GTGAGGCTCCCACAGCACCAGGAGCACCAGCAGCCTGAATGTAGCCCTTCCAACACCTGAC   1920

Qy     1926  AGAATGAGAACCAGCAGCACCAGCAGCCAAACCAAGACCATCTGAGGAATG   1985
             |||||||||||||||||||||||||||||||||||||||||||||||||
Db     1921  AGAATGAGAACCAGCAGCACCAGCAGCCAAACCAAGACCATCTGAGGAATG   1980

Qy     1986  AGAAGTCTGCGGAGGCGGCGGCTCTGCCGAGGCCCTGAGAACCCCAGGGGCCGAGG   2045
             |||||||||||||||||||||||||||||||||||||||||||||||||
Db     1981  AGAAGTCTGCGGAGGCGGCGGCTCTGCCGAGGCCCTGAGAACCCCAGGGGCCGAGG   2040
```

Qy 2046 AGGGGCGGGGAAGGTCAGCCAGGTTGCCAGAACCACCGAGCCGCCTTCCCGCCCCCA 2105
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2041 AGGGGCGGGGAAGGTCAGCCAGGTTGCCAGAACCACCGAGCCGCCTTCCCGCCCCCA 2100

Qy 2106 GGGCTTCTGCGAGGCTTCAGCCATCCACTTCACCATCCGGATCTCTCCTGAACTCCC 2165
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2101 GGGCTTCTGCGAGGCTTCAGCCATCCACTTCACCATCCGGATCTCTCCTGAACTCCC 2160

Qy 2166 ACGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAGCCAAGCAAAAT 2225
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2161 ACGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAGCCAAGCAAAAT 2220

Qy 2226 GCACACCCTTTTCTGTGGCAAATCGTCTCGTACACAGGTGCCTGCAGCGGTAATATTTAG 2285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2221 GCACACCCTTTTCTGTGGCAAATCGTCTCGTACACAGGTGCCTGCAGCGGTAATATTTAG 2280

Qy 2286 GATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGGTAATATTTTTAG 2345
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2281 GATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGGTAATATTTTTAG 2340

Qy 2346 AAATATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTAATAATTTTTTTTT 2405
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2341 AAATATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTAATAATTTTTTTTT 2400

Qy 2406 CTTTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAAGATAAAAGTCTTTGGGAGGTCTC 2465
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2401 CTTTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAAGATAAAAGTCTTTGGGAGGTCTC 2460

Qy 2466 ACGGTGTAGAGAGGAGGAGCTTTGAGGCCACCCGCACAAAAATTCACCCAGAGGGAAATCTCGT 2525
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2461 ACGGTGTAGAGAGGAGGAGCTTTGAGGCCACCCGCACAAAAATTCACCCAGAGGGAAATCTCGT 2520

Qy 2526 CGGAAGGACACTCACGGCAGTTCTGATCACCTGTGTATGTCAACAGAAGGGATACCGTC 2585
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2521 CGGAAGGACACTCACGGCAGTTCTGATCACCTGTGTATGTCAACAGAAGGGATACCGTC 2580

Qy 2586 TCCTTGAAGAGGAAACTCTGTCACTCCTGTCAGCTCATCACACCCATTTCTCT 2645
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2581 TCCTTGAAGAGGAAACTCTGTCACTCCTGTCAGCTCATCACACCCATTTCTCT 2640

Qy 2646 TTGCTTCACAGGTTTTCAAACTGGTTTTTTTGCATACTGCTATATAATTCTCTGTCTCTCTC 2705
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2641 TTGCTTCACAGGTTTTTTAAACTGGTTTTTTTGCATACTGCTATATAATTCTCTGTCTCTCTC 2700

Qy 2706 TGTTTATCTCTCCCCCTCCCCCTCCCCCCTTCTTCTCCATCTCCATTCTTTTGAATTTCC 2765
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2701 TGTTTATCTCTCCCCCTCCCCCTCCCCCCTTCTTCTCCATCTCCATTCTTTTGAATTTCC 2760

Qy 2766 TCATCCCTCCATCTCAATCCCGTATCGTATCCCGTATCTACGGCA-CCCCCCCCCCCAGGCAGTGCT 2825
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2761 TCATCCCTCCATCTCAATCCCGTATCTCCGTATCTACGGCA-CCCCCCCCCCCAGGCAGTGCT 2819

Qy 2826 CTGAGTATCACATCACACACAAAAGGAACAAAAGCGAAACACACAACAACAGAGATTCC 2885
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2820 CTGAGTATCACATCACACACAAAAGGAACAAAAGCGAAACACAACAACAGAGATTCC 2879

Qy 2886 ACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGGAA 2945
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2880 ACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGGAA 2939

Qy 2946 AACAGGAACCACCACCAACCAATCAACCAAACAAAGAAAAATTCCAACAATGAAAGA 3005
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2940 AACAGGAACCACCACCAACCAATCAACCAAACAAAGAAAAATTCCAACAATGAAAGA 2999

Qy 3006 ATGTATTTTGTCTTTTGTCATTTGCATTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCC 3065
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3000 ATGTATTTTGTCTTTTGTCATTTGCATTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCC 3059

Qy 3066 TTTCTTT-AAAAAAAAAAAAAAAAAAAGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGGCCCAGG 3124
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3060 TTTCTTTAAAAAAAAAAAAAAAAAAAAAGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGGCCCAGG 3119

Qy 3125 GCGTTAAATTCACAGATTTTTTAACGAGAAAAACACAGAAGAGAAGCTACCTCAGGTGT 3184
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3120 GCGTTAAATTCACAGATTTTTTAACGAGAAAAACACAGAAGAGAAGCTACCTCAGGTGT 3179

Qy 3185 TTTTACCCTCAGCACCTTGCTCTTGTGTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGG 3244
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3180 TTTTACCCTCAGCACCTTGCTCTTGTGTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGG 3239

Qy 3245 AGCATTTTTTATTTTTTAATAAAAATGAGTGGAAAAAAATAAGATAATCAACTGCCA 3304
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3240 AGCATTTTTTATTTTTTAATAAAAATGAGTGGAAAAAAATAAGATAATCAACTGCCA 3299

Qy 3305 GCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCTAGCC 3364
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3300 GCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCTAGCC 3359

Qy 3365 AAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3412
          ||||||||||||||||||||||||||||||||||||||||||||||
Db 3360 AAGAACCTATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3407

RESULT 5
AR171867
LOCUS   AR171867   3283 bp   DNA   linear   PAT 17-DEC-2001
DEFINITION Sequence 8 from patent US 6297364.
ACCESSION   AR171867
VERSION   AR171867.1 GI:17910817
KEYWORDS   .
SOURCE   Unknown.
  ORGANISM Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 3283)
  AUTHORS   Chen,Y.-T., Gure,A., Tsang,S., Stockert,E., Jager,E., Alexander,K.
            and Old,L.J.
  TITLE   Isolated nucleic acid molecule encoding cancer associated antigen,
            the antigen itself, and uses thereof
  JOURNAL   Patent: US 6297364-A 8 02-OCT-2001;
FEATURES   Location/Qualifiers
     source   1..3283
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Query Match   92.1%; Score 3143; DB 6; Length 3283;
Best Local Similarity 96.2%; Pred. No. 0;
Matches 3283; Conservative   0; Mismatches   0; Indels   129; Gaps   1;

Qy   1 GGCAGCGGAGGAGGAGGAGGCGCCGGGTACCGGCGCGGGGGGAGCGGGTACCGGGCGGGGCGGCTCTCGGGG 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGCAGCGGAGGAGGAGGAGGCGCCGGGTACCGGCGCGGGGGGAGCGGGTACCGGGCGGGGCGGCTCTCGGGG 60

Qy   61 AAGAGACGGATGATGAACAAGCTTTACATCGGAAGCTTTACATCGGAAGCTGAGCCCCGTCACCGCCGAC 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AAGAGACGGATGATGAACAAGCTTTACATCGGAAGCTTTACATCGGAAGCTGAGCCCCGTCACCGCCGAC 120

Qy 121 GACCTCCGCCAGCTCTTTGGGACAGGAAGCTGCCCTGCGGGACAGGTCCTGCTGAAG 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 GACCTCCGCCAGCTCTTTGGGACAGGAAGCTGCCCTGCGGGACAGGTCCTGCTGAAG 180

Qy 181 TCCGGCTACGCCTTCGTGCTGGGACTACCCCGACTACCCCGACCAGGAACTGGGCCATCCGGCCATCCGAGACC 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 TCCGGCTACGCCTTCGTGCTGGGACTACCCCGACTACCCCGACCAGGAACTGGGCCATCCGGCCATCCGAGACC 240

Qy 241 CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA 300

Qy 301 AAGCTAAGGAGCAGGAAAATTCAGATTCGAAACATCCCTCCTACCTCAGTGGAGGTG 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 AAGCTAAGGAGCAGGAAAATTCAGATTCGAAACATCCCTCCTACCTCAGTGGAGGTG 360

Qy 361 TTGGATGGACTTTTGGCTCAATATGGACGATGGAATGTGGACAACAAGTGAACAGACAGAC 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 361 TTGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGGACAAGTCAACACAGAC 420

Qy 421 ACAGAAACCGCCGTGTCAACGTCACATATGCAACAAGAGAAGCAAAAAATAGCCATG 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 ACAGAAACCGCCGTGTCAACGTCACATATGCAACAAGAGAAGCAAAAAATAGCCATG 480

Qy 481 GAGAAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGAT 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 GAGAAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGAT 540

Qy 541 GAAGAGGTGAGCTCCCCTTCCCCCCTCAGCGAGCCCAGCGTGGGGACCACTCTTCCCG 600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 GAAGAGGTGAGCTCCCCTTGCCCCCTCAGCGAGCCCAGCGTGGGGACCACTCTTCCCG 600

Qy 601 GAGCAAGGCCACGCCCCTGGGGGCACTTCTCAGGCCACAGATTGATTTCCCGCTGCGG 660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 GAGCAAGGCCACGCCCCTGGGGGCACTTCTCAGGCCACAGATTGATTTCCCGCTGCGG 660

Qy 661 ATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGAGGGGCTTGACCATAAAG 720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 661 ATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGAGGGGCTTGACCATAAAG 720

Qy 721 AACATCACTAAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT 780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 721 AACATCACTAAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT 780

Qy 781 GCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATG 840
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 781 GCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATG 840

Qy 841 ATTCTTGAAATCATGCAGAAAGAGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG 900
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 841 ATTCTTGAAATCATGCAGAAAGAGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG 900

Qy 901 AAAATCTTGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAAATTTG 960
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 901 AAAATCTTGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAAATTTG 960

Qy 961 AAGAAAATTGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 961 AAGAAAATTGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020

Qy 1021 ATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCGTGCCAGTGCT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 ATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCGTGCCAGTGCT 1080

Qy 1081 GAGATAGAGATTATGAAGAAGTGCGTGAGGGCCTTTGAAAATGATATGCTGGCTGTTAAC 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 GAGATAGAGATTATGAAGAAGTGCGTGAGGGCCTTTGAAAATGATATGCTGGCTGTTAAC 1140

Qy 1141 CAACAAGCCAATCTCGATCCCAGGGTTGAACCTCAGCGCCACTTGCCATCTTTTCAAACAGGA 1200

Db 1141 -- 1140

Qy 1201 CTGTCCGTGCTATCTCCACCAGCGGGCCCCCGCGGAGCTCCCCCGTGCCCCCTACCAC 1260

Db 1141 -- 1140

Qy 1261 CCCTTCACTCACCTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTTGGCCCCG 1320
          ||||||||||||||||||||||||||||||||
Db 1141 ------ACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTTGGCCCCG 1191

Qy 1321 TTTCCCGCATCATCACTCTTCATCGGGAAGACAGAGGAGGGGCACACATCAAACAGCTGGC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1192 TTTCCCGCATCATCACTCTTCATCGGGAAGACAGAGGAGGGGCACACATCAAACAGCTGGC 1251

Qy 1381 GCTGTGGGCGCCATCATCGGGAAGAAGGGGCACACATCAAACAGCTGGCGAAGATTCGCC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1252 GCTGTGGGCGCCATCATCGGGAAGAAGGGGCACACATCAAACAGCTGGCGAAGATTCGCC 1311

Qy 1441 GGAGCCTCTATCAAGATTGCCCCTGCCGAAGGCCAGACGTCAGCGAAAGGATGGTCATC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1312 GGAGCCTCTATCAAGATTGCCCCTGCCGAAGGCCAGACGTCAGCGAAAGGATGGTCATC 1371

Qy 1501 ATCACCGGCCACCGGAAGCCCAGTTGCAAGGCCCAGGACGGACGGATCTTTGGGAAAACTGAAA 1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1372 ATCACCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGACGGACGGATCTTTGGGAAAACTGAAA 1431

Qy 1561 GAGGAAAACTTCTTTAACCCCAAAGAGAAGAAGTGAAGCTGGAAGCGCCATATCAGAGTGCCC 1620
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1432 GAGGAAAACTTCTTTAACCCCAAAGAGAAGAAGTGAAGCTGGAAGCGCCATATCAGAGTGCCC 1491

Qy 1621 TCTTCCACAGCTGGCCGGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTGCAGAAC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1492 TCTTCCACAGCTGGCCGGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTGCAGAAC 1551

Qy 1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGCCACAAACGCCAGATGAAAATGAGGAAGTG 1740
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1552 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGCCACAAACGCCAGATGAAAATGAGGAAGTG 1611

Qy 1741 ATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1612 ATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAA 1671

Qy 1801 ATTGTACAACAGGTGAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1672 ATTGTACAACAGGTGAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGC 1731

Qy 1861 AGCAAGTGAGGGTCCACAGGCACCAGCAAACAACGGATGAATGAATGTAGCCCTTCCAACAC 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1732 AGCAAGTGAGGGTCCCACAGCACCAGCAAAACAACGGATGAATGAATGTAGCCCTTCCAACAC 1791

Qy 1921 CTGACAATGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAG 1980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1792 CTGACAATGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAG 1851

Qy 1981 GAATGAGAAGTCTGCGGAGGCGGCCAGGTTCAGCGAGGCTTCAGCCTTCCAGAACCACCCCAGGGGC 2040
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1852 GAATGAGAAGTCTGCGGAGGCGGCCAGGTTCAGCGAGGCTTCAGCCTTCCAGAACCACCCCAGGGGC 1911

Qy 2041 CGGAGGAGGGGCGGGGAAGGTCAGCCAGGTTTGCCAGAAGAAGGCAGATTCTTGAGACCCCGCCTCCCGCGC 2100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1912 CGGAGGAGGGGCGGGGAAGGTCAGCCAGGTTTGCCAGAAGAAGGCAGATTCTTGAGACCCCGCCTCCCGCC 1971

Qy 2101 CCCCAGGGCTTCTGCAGGCTTCGAGGCTTCAGCGCTATCCTTTTAGTTGAACTAACAGGTGAAGCTGTTCAAAGCCAAGC 2160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1972 CCCCAGGGCTTCTGCAGGCTTCGAGGCTTCAGCGCTATCCTTTTAGTTGAACTAACAGGTGAAGCTGTTCAAAGCCAAGC 2031

Qy 2161 CTCCCACGACGCTATCCCTTTTAGTTGAACTAACAGGTGAAGCTGTTCAAAGCCAAGC 2220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2032 CTCCCACGACGCTATCCCTTTTAGTTGAACTAACAGGTGAAGCTGTTCAAAGCCAAGC 2091

Qy 2221 AAAATGCACACCCTTTTTCTGTGGCAAATCGTCTCTGTACATATTAGAAA 2280
          ||||||||||||||||||||||||||||||||||||||||||
Db 2092 AAAATGCACACCCTTTTTCTGTGGCAAATCGTCTCTGTACATATTAGAAA 2151

Qy 2281 GGGAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGGTAATATAT 2340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2152 GGGAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGGTAATATAT 2211

Qy 2341 TTTAGAAATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTT 2400
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2212 TTTAGAAATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTT 2271

Qy 2401 TTTTTCTTTTAAAGAGAAAGAAAGCCTTTGAGCGTAGACTTTAAAGAATAAAGTCTTTGGGAG 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2272 TTTTTCTTTTAAAGAGAAAGAAAGCAGGCTTTGAGCGTAGACTTTAAAGAATAAAGTCTTTGGGAG 2331

Qy 2461 GTCTCACGGTGTAGAGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAGGGAAAT 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2332 GTCTCACGGTGTAGAGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAGGGAAAT 2391

Qy 2521 CTCGTCGGAAGGACGTCCCCTGCCGAAGGCCAGACGTCAGCGAAAGGATGTCATC 2580
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2392 CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAGGGATA 2451

```
FEATURES          Location/Qualifiers
     source          1..3283
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Query Match          92.1%;  Score 3143;  DB 6;  Length 3283;
Best Local Similarity  96.2%;  Pred. No. 0;
Matches 3283;  Conservative  0;  Mismatches  0;  Indels 129;  Gaps  1;

Qy     1 GGCAGCGGAGGAGGCGGAGGAGCGGCCGGGTACCGGCGGGGAGGGGTACCGGCCTCGGGG 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GGCAGCGGAGGAGGCGGAGGAGCGGCCGGGTACCGGCGGGGAGGGGTACCGGCCTCGGGG 60

Qy    61 AAGAGACGGATGACAAGCTTACATCGGGAAGCTTGAGCCCGTCACCGGCCGAC 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 AAGAGACGGATGACAAGCTTACATCGGGAAGCTTGAGCCCGTCACCGGCCGAC 120

Qy   121 GACCTCCGGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGGCGGGACAGGTCCTGCTGGAAG 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GACCTCCGGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGGCGGGACAGGTCCTGCTGGAAG 180

Qy   181 TCCGGCTACGCCTTCGTGGACTACCCGACCAGAACTGGCCATCCGCCGCCATCCGAGACC 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TCCGGCTACGCCTTCGTGGACTACCCGACCAGAACTGGCCATCCGCCGCCATCCGAGACC 240

Qy   241 CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGGAAAATCAGTCTCTAA 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGGAAAATCAGTCTCTAA 300

Qy   301 AAGCTAAGGAGCAGGAAAATTCAGATTCGAAACATCCCTGCAGTGGAGAGTG 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AAGCTAAGGAGCAGGAAAATTCAGATTCGAAACATCCCTGCAGTGGAGAGTG 360

Qy   361 TTGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGAC 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TTGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGAC 420

Qy   421 ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAGAGAAGCAAAATAGCCATG 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAGAGAAGCAAAATAGCCATG 480

Qy   481 GAGAAGCTAAGCGGCATCAGTTGAGAACTACTCCTTCCTTCAAGATTTCCTACATCCCGGAT 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GAGAAGCTAAGCGGCATCAGTTGAGAACTACTCCTTCCTTCAAGATTTCCTACATCCCGGAT 540

Qy   541 GAAGAGGTGAGCTCCTTCGCCCCTCAGCGTGGGGACCACTCTTCCCG 600
         |||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GAAGAGGTGAGCTCCTTCGCCCCTCAGCGTGGGGACCACTCTTCCCG 600

Qy   601 GAGCAAGGCCACGCCCCTGGGGGCACTTCTCAGGCAGCAGATTGATTTCCCGCTGCGG 660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GAGCAAGGCCACGCCCCTGGGGGCACTTCTCAGGCAGCAGATTGATTTCCCGCTGCGG 660

Qy   661 ATCCTGTCCCCACCCAGTTGTTGTGGTGCCATCATCGGAAAAGAGAACTCTGGAGCT 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 ATCCTGTCCCCACCCAGTTGTTGTGGTGCCATCATCGGAAAAGAGAACTCTGGAGCT 720

Qy   721 AACATCACTAAGCACAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT 780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 AACATCACTAAGCACAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT 780

Qy   781 GCAGAGAAGCCTGTCACCATCGCACCCCAGGAGGGACTTCTGAAGCCATGCGCATG 840
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GCAGAGAAGCCTGTCACCATCGCACCCCAGGAGGGACTTCTGAAGCCATGCGCATG 840

Qy   841 ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG 900
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG 900

Qy   901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTG 960
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTG 960

Qy  2581 CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTCATGCCTGTCTCAGCTCATACACCCATT 2640
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2452 CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTCATGCCTGTCTCAGCTCATACACCCATT 2511

Qy  2641 TCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTTTGCATATACTGCTATATAATCTCTGTCT 2700
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2512 TCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTTTGCATATACTGCTATATAATCTCTGTCT 2571

Qy  2701 CTCTCGTGTTTATCTCTCCCCCTCCCCCCTTCTTCTTCCATCTCCATTCTTTGAA 2760
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2572 CTCTCGTGTTTATCTCTCCCCCTCCCCCCTTCTTCTTCCATCTCCATTCTTTGAA 2631

Qy  2761 TTTCCTCATCCCTCCATCTCAATCCGTATCTACGCACCCCGTATCCGTATCTACGCAAAGCA 2820
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2632 TTTCCTCATCCCTCCATCTCAATCCGTATCTACGCACCCCGTATCCGTATCTACGCAAAGCA 2691

Qy  2821 GTGCTCTGAGTATCACATCACACAAAGGAACAAAAGCGAAACACACACCAGACCCTCAA 2880
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2692 GTGCTCTGAGTATCACATCACACAAAGGAACAAAAGCGAAACACACACCAGACCCTCAA 2751

Qy  2881 CTTACACTTGGTTACTCAAAAGAACAGAGTCAATGGTACTTGTCTCCTAGCGTTTTGGAAG 2940
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2752 CTTACACTTGGTTACTCAAAAGAACAGAGTCAATGGTACTTGTCTCCTAGCGTTTTGGAAG 2811

Qy  2941 AGGAAAACAGGAACCACCAACAATCAACCAACCAACAAAGAAAAATTCCAACAATG 3000
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2812 AGGAAAACAGGAACCACCAACAATCAACCAACCAACAAAGAAAAATTCCAACAATG 2871

Qy  3001 AAAGAATGTATTTTGTCTTTTGCATTTTGGTGTATAAGCCATCAATATTCAGCAAAATG 3060
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2872 AAAGAATGTATTTTGTCTTTTGCATTTTGGTGTATAAGCCATCAATATTCAGCAAAATG 2931

Qy  3061 ATTCCTTTCTTAAAAAAAAAAATGTGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCC 3120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2932 ATTCCTTTCTTAAAAAAAAAAATGTGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCC 2991

Qy  3121 CAGGGCGTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAAGAAGCTACCTCAG 3180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2992 CAGGGCGTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAAGAAGCTACCTCAG 3051

Qy  3181 GTGTTTTTACCTCAGCACCTTGCTCTTGTCTTGTGTTCCCTTAGAGATTTGTAAAGTGATAG 3240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3052 GTGTTTTTACCTCAGCACCTTGCTCTTGTCTTGTGTTCCCTTAGAGATTTGTAAAGTGATAG 3111

Qy  3241 TTGGAGCATTTTTTATTTTTAATAAAAATGAGTTGAAAAAAATAAGAATATCAACT 3300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3112 TTGGAGCATTTTTTATTTTTAATAAAAATGAGTTGAAAAAAATAAGAATATCAACT 3171

Qy  3301 GCCAGCCTGGAGAAGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCT 3360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3172 GCCAGCCTGGAGAAGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCT 3231

Qy  3361 AGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3412
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3232 AGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3283
```

RESULT 6

```
AR343076
LOCUS       AR343076                3283 bp    DNA     linear   PAT 17-AUG-2003
DEFINITION  Sequence 8 from patent US 6576755.
ACCESSION   AR343076
VERSION     AR343076.1  GI:33738478
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 3283)
  AUTHORS   Chen,Y.-T., Gure,A., Tsang,S., Stockert,E., Jager,E., Alexander,K.
            and Old,L.J.
  TITLE     Isolated nucleic acid molecule encoding cancer associated antigen,
            the antigen itself, and uses thereof
  JOURNAL   Patent: US 6576755-A 8 10-JUN-2003;
```

```
Qy   961  AAGAAAAATTGAACATGAAAACAGGGACCAAGATAACAATTCATCTTTGCAGGATTTGAGC  1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961  AAGAAAAATTGAACATGAAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC  1020

Qy  1021  ATATACAAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCT  1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021  ATATACAAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCT  1080

Qy  1081  GAGATAGAGATTATGAAGAAGCTGCGTGAGCGCTTTGAAAATGATATGCTGGTGTAAC  1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081  GAGATAGAGATTATGAAGAAGCTGCGTGAGCGCTTTGAAAATGATATGCTGGCTGTTAAC  1140

Qy  1141  CAACAAGCCAATCTGATCCCAGGGTTGAACCTGCCACTTGGCATCTTTTCAACAGGA  1200
          ------------------------------------------------------------
Db  1141                                                               1140

Qy  1201  CTGTCCGTGCTATCTCCACCAGCAGGGCCCCGCGGAGCTCCCCCGCTGCCCCCTACCAC  1260
          ------------------------------------------------------------
Db  1141  -----                                                          1140

Qy  1261  CCCTTCACTACCACTCCGGATACTTCTCCAGCCATCACCCCATCACCAGTTTGGGCCCG  1320
          ------                          |||||||||||||||||||||||||||||
Db  1141  ------ACCCACTCCGGATACTTCTCCAGCCATCACCCCATCACCAGTTTGGGCCCG  1191

Qy  1321  TTCCCGCCATCATCATCTTATCCAGAGCAGGAGATTGTGAATCTCTTCATCCCAACCCAG  1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1192  TTCCCGCCATCATCATCTCATTATCCAGAGCAGGAGATTGTGAATCTCTTCATCCCAACCCAG  1251

Qy  1381  GCTGTGTGGGGCGCCATCATCGGGAAGAAGGGGGCCACACATCAAAACAGTCGGCGAGATTCGCC  1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1252  GCTGTGTGGGGCGCCATCATCGGGAAGAAGGGGGCCACACATCAAACAAAAT CGGCGAGATTCGCC  1311

Qy  1441  GGAGCCTCTATCAAGATTGCCCCTGGCAAGAAGTGAAGCTGGAGCGCCATATCAGAGTGCCC  1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1312  GGAGCCTCTATCAAGATTGCCCCTGGCAAGAAGTGAAGCTGGAGCGCCATATCAGAGTGCCC  1371

Qy  1501  ATCACCGGGCCACCCGGAAGCCAGTTCAAGGCCCAGGACGGATCTTTGGGAAACTGAAA  1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1372  ATCACCGGGCCACCCGGAAGCCAGTTCAAGGCCCAGGACGGATCTTTGGGAAACTGAAA  1431

Qy  1561  GAGGAAAACTTCTTTAACCCCAAAGAAGTGAAGCTGGAGCGCATATCAGAGTGCCC  1620
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1432  GAGGAAAACTTCTTTAACCCCAAAGAAGAGAGTGAAGCTGGAGCGCATATCAGAGTGCCC  1491

Qy  1621  TCTTCCACAGCTGGCGGGTGATTGGCAAGAGTGCAAGACCGTGCAAGCGAACTGCAGAAC  1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1492  TCTTCCACAGCTGGCGGGTGATTGGCAAGAGGTGCAAGACCGTGCAAGACCGTGCAAGAAC  1551

Qy  1681  TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGAGAAATGAGGGAAGTG  1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1552  TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGAGAAATGAGGGAAGTG  1611

Qy  1741  ATCGTCAGAATTATCGGGCACTTCTTTGCTCTGCCTCGGACTAGCCTTCACAGCGCGC  1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1612  ATCGTCAGAATTATCGGGCACTTCTTTGCTCTGCCTCGGACTAGCCTTCACAGCGCGC  1671

Qy  1801  ATTGTACACAGGTTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTGCCTTCCAACGCGC  1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1672  ATTGTACACAGGTTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTGCCTTCCAACGCGC  1731

Qy  1861  AGCAAGTGAGGCTCTGCAGGCCACCAGCAGACGCAGCAGGCCGCAAAAGACCATCTGAG  1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1732  AGCAAGTGAGGCTCTGCAGGCCACCAGCAGACGCAGCAGGCCGCAAACAGACCATCTGAG  1791

Qy  1921  CTGACAGAATGAGACACGAACTGCCGAGGGCGGCCAGGCCCTCCAACAC  1980
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  1792  CTGACAGAATGAGACACGAACTGCCGAGGGCGGCCAGGCCCTCCAACAC  1851

Qy  1981  GAATGAGAAGTCTGCGCAGGGCGGCCGGCCCAGGGGACTCTGCCCGAGGGGGC  2040
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1852  GAATGAGAAGTCTGCGCAGGGCGGCCGGCCCAGGGGACTCTGCCCGAGGGGGC  1911
```

```
Qy  2041  CGAGGAGGGCGGGGGGAAGTCAGCCAGGTTGCCAGAACCACCGAGCCCGCCTCCCGCC  2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1912  CGAGGAGGGCGGGGGGAAGTCAGCCAGGTTGCCAGAACCACCGAGCCCGCCTCCCGCC  1971

Qy  2101  CCCCAGGGCTTCTGCAGCCTTCAGCCATCCAGCCATCCACCATCCACTCCGGATCTCTCCTGAA  2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1972  CCCCAGGGCTTCTGCAGCCTTCAGCCATCCAGCCATCCACCATCCACTCCGGATCTCTCCTGAA  2031

Qy  2161  CTCCCACGACGCTATCCCTTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGC  2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2032  CTCCCACGACGCTATCCCTTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGC  2091

Qy  2221  AAAATGCACACCCTTTTCTGTGCAAATCTCCTCTGTACTGTGTGTGTACTATTAGAAA  2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2092  AAAATGCACACCCTTTTCTGTGCAAATCGTCTCTGTACTGTGTGTACTATTAGAAA  2151

Qy  2281  GGGAAGATGTTAAGATATGTGGCCTGGTGGTACACAGGGTGCCTGCAGCCGGTAATATAT  2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2152  GGGAAGATGTTAAGATATGTGGCCTGGTGGTACACAGGGTGCCTGCAGCCGGTAATATAT  2211

Qy  2341  TTTAGAAATAATATATCAAATAACTCCAATTTTAATCAATTATTAATTTT  2400
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2212  TTTAGAAATAATATATCAAATAACTCCAATTTTAATCAATTATTAATTTT  2271

Qy  2401  TTTTTCTTTTAAAGAGAAAGCAGGCTTTCTTCTTAAAGAATAAAGTCTTTGGGAG  2460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2272  TTTTTCTTTTAAAGAGAAAGCAGGCTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG  2331

Qy  2461  GTCTCACGTGTGAGAGAGGAGCTTTGAGGCGCCACCCGACAAAATTCACCCAGAGGGAAAT  2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2332  GTCTCACGTGTGAGAGAGGAGCTTTGAGGCGCCACCCGACAAAATTCACCCAGAGGGAAAT  2391

Qy  2521  CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCGTGTATGTCAACGAGAGGGATA  2580
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2392  CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCGTGTATGTCAACGAGAGGGATA  2451

Qy  2581  CCGTCTCCTTGAAGAGAAACTCTGTCACTCCTGTCTCAGCTCATCACCATT  2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2452  CCGTCTCCTTGAAGAGAAACTCTGTCACTCCTGTCTCAGCTCATCACCATT  2511

Qy  2641  TCTCTTTGCTTCACAGTTTTTAAACTGGTTTTTTGCATACTGCTATATAAATTCTCTGTCT  2700
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2512  TCTCTTTGCTTCACAGTTTTTAAACTGGTTTTTTGCATACTGCTATATAAATTCTCTGTCT  2571

Qy  2701  CTCTCTGTTTTATCTCTCCCCTCCCCTCCCCTCCATTCTCATTCTTTGAA  2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2572  CTCTCTGTTTTATCTCTCCCCTCCCCTCCCCTCCCATTCTCATTCTTTGAA  2631

Qy  2761  TTTCCTCATCCCTCCATCTCAATCCTCTCAATCCGTATCCTCACGCACCCCCCAGGCAAAGCA  2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2632  TTTCCTCATCCCTCCATCTCAATCCTCTCAATCCGTATCCTCACGCACCCCCCAGGCAAAGCA  2691

Qy  2821  GTGCTCGAGTATCACATCACACAAGGAACAACAAGAGTCAATGGTACTTGTCCTAGGGAAG  2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2692  GTGCTCGAGTATCACATCACACAAGGAACAACAAGAGTCAATGGTACTTGTCCTAGGGAAG  2751

Qy  2881  CTTACACTTGGTTACTCAACAAGGACAAGACAACAACCAACAAAGAAAAAATTCCACCAATG  2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2752  CTTACACTTGGTTACTCAACAAGGACAAGACAACAACCAACAAAGAAAAAATTCCACCAATG  2811

Qy  2941  AGGAAAACAGGAACCCACCCAAACCAAATCAACCACCAACAAAGAAAAATTCCACAATG  3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2812  AGGAAAACAGGAACCCACCCAAACCAAATCAACCACCAACAAAGAAAAATTCCACAATG  2871

Qy  3001  AAAGAATGTATTTTGTCTTTTGCATTTTGGTGTATAAGCCATCAGCAAAATG  3060
          ||||||||||||||||||||||||||||||||||||||||||||||
Db  2872  AAAGAATGTATTTTGTCTTTTGCATTTTGGTGTATAAGCCATCAGCAAAATG  2931

Qy  3061  ATTCCTTTCTTTAAAAAAAAATGTGGAGAAAGTAGAAATTTACCAAGGTTGTTGGCC  3120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2932  ATTCCTTTCTTTAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGGC  2991

Qy  3121  CAGGGCGTTAAATTCACAGATTTTTTTAACGAGAAAAACACACAGAAGAGCTACCTCAG  3180
```

Db   2992 CAGGGCGTTAAATTCACAGATTTTTTAACGAGAGAAAACACACAGAGAAGAAGCTACCTCAG 3051
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   3181 GTGTTTTTACCTCAGCACCTTGCTCTTGTTCCCTTAGAGATTTTGTAAAGCTGATAG 3240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3052 GTGTTTTTACCTCAGCACCTTGCTCTTGTTCCCTTAGAGATTTTGTAAAGCTGATAG 3111

Qy   3241 TTGGAGCATTTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACT 3300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3112 TTGGAGCATTTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACT 3171

Qy   3301 GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCT 3360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3172 GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCT 3231

Qy   3361 AGCCAAGAGACCNATATGGCCTTCTTTGGTGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3412
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3232 AGCCAAGAGACCNATATGGCCTTCTTTGGTGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3283

RESULT 7
BD209927
LOCUS   BD209927   3283 bp   DNA   linear   PAT 17-JUL-2003
DEFINITION Isolated nucleic acid molecules encoding cancer-associated
            antigens, these antigens and method of using the same.
ACCESSION   BD209927
VERSION   BD209927.1 GI:33019697
KEYWORDS   JP 2002512049-A/6.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3283)
  AUTHORS   Chen,Y.T., Gure,A., Tsang,S., Stockert,E., Jager,E., Knuth,A. and
            Old,L.J.
  TITLE   Isolated nucleic acid molecules encoding cancer-associated
            antigens, these antigens and method of using the same
  JOURNAL   Patent: JP 2002512049-A 6 23-APR-2002;
            LUDWIG INSTITUTE FOR CANCER RESEARCH
COMMENT   OS   Homo sapiens (human)
            PN   JP 2002512049-A/6
            PD   23-APR-2002
            PF   16-MAR-1999 JP 2000545030
            PR   17-APR-1998 US 09/061709
            PI   YAO TSENG CHEN,ALI GURE,SOLAM TSANG,ELISABETH STOCKERT,ELKE
            JAGER,
            PI   ALEXANDER KNUTH,LLOYD J OLD
            PC   C12N15/09,A61K35/12,A61K39/00,A61P35/00,C07K16/32,
            PC   C12N1/15,
            PC   C12N1/19,C12N1/21,C12N5/10,C12P21/08,C12Q1/68,G01N33/53,G01N33/ PC
            574,
            PC   C12N15/00,C12N5/00
            CC   n=not determined(3243)
FH   Key   Location/Qualifiers
FT   source   1..3283
FT   /organism='Homo sapiens (human)'.
            source   1..3283
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"

ORIGIN

Query Match   92.1%; Score 3143; DB 6; Length 3283;
Best Local Similarity 96.2%; Pred. No. 0;
Matches 3283; Conservative   0; Mismatches   0; Indels 129; Gaps   1;

Qy   1 GGCAGCGGAGGAGGCGAGGGCGAGAGCGCCGGGTACCGGGAGCCCGGGGAGCCCGGGGAGGCTCTCGGGGG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGCAGCGGAGGAGGCGAGGGCGAGAGCGCCGGGTACCGGGAGCCCGGGGAGCCCGGGGAGGCTCTCGGGGG 60

Qy   61 AAGAGACGGATGATGAACAAGCTTTACATCGGGAACCTTGAGCCCGTCGTCACCGCCGGAC 120

```
Db  1141 ------------------------------------------------------------ 1140

Qy  1201 CTGTCCGTGCTATCTCCACCAGCAGGCCCCGCGGAGCTCCCCCGCTCCCCCCTACCAC 1260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 ------------------------------------------------------------ 1140

Qy  1261 CCCTTCACTACCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTGGCCCG 1320
                   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 ------ACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTGGCCCG 1191

Qy  1321 TTCCCGCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTTCTTCATCCCAACCCAG 1380
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1192 TTCCCCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTTCTTCATCCCAACCCAG 1251

Qy  1381 GCTGTGGGCGCCATCATCGGGAAGAAGAGGGGCACACATCAAAACAGCTGCGAGATTCGCC 1440
         ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db  1252 GCTGTGGGCGCCATCATCGGGAAGAAGAGGGGCACCATCAAAACAGCTGCGAGATTCGCC 1311

Qy  1441 GGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATC 1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1312 GGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATC 1371

Qy  1501 ATCACCGGGCACCGGAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAA 1560
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1372 ATCACCGGGCACCGGAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAA 1431

Qy  1561 GAGGAAAACTTCTTTAACCCCAAAGAAGAGTGAAGCTGAAGGCATATCAGAGTGCCC 1620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1432 GAGGAAAACTTCTTTAACCCCAAAGAAGAGTGAAGCTGAAGGCATATCAGAGTGCCC 1491

Qy  1621 TCTTCCACAGCTGCCGGGTGATTGGCAAGGTGGCAAGACCGTGACGAACTGCAGAAC 1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1492 TCTTCCACAGCTGCCGGGTGATTGGCAAGGTGGCAAGACCGTGACGAACTGCAGAAC 1551

Qy  1681 TTAACCAGTTGCAGAATTATCGGGCACTTCTTTGCTAGCCAGAGAAATACCCTCAGGGAGTG 1740
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1552 TTAACCAGTTGCAGAATTATCGGGCACTTCTTTGCTAGCCAGAGAAATACCCTCAGGGAGTG 1611

Qy  1741 ATCGTCAGAATTATCGGGCACTGTGCCCTCGTGACCAAAGCGCGTGACAGCGAAGATCAGGGAA 1800
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1612 ATCGTCAGAATTATCGGGCACTGTGCCCTCGTGACCAAAGCGCGTGACAGCGAAGATCAGGGAA 1671

Qy  1801 ATTGTACAACAGGTGAAGCAGCAGGAGCAGAGAAATACCCTCAGGGAGTGCCTTCACAGCGC 1860
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1672 ATTGTACAACAGGTGAAGCAGCAGGAGCAGAGAAATACCCTCAGGGAGTGCCTTCACAGCGC 1731

Qy  1861 AGCAAGTGAGGCGGGGGAAGTCAGCCAGGTCAGCCAGGGGCCTGCCGAGGGCCTGAGAATG 1920
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1732 AGCAAGTGAGGCGGGGGAAGTCAGCCAGGTCAGCCAGGGGCCTGCCGAGGGCCTGAGAATG 1791

Qy  1921 CTGACAGAATGAGACAAAACGCAGCCAGCCAGATCGGAGGCAAAACCAAAGACCATCTGAG 1980
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1792 CTGACAGAATGAGACAAAACGCAGCCAGCCAGATCGGAGGCAAAACCAAAGACCATCTGAG 1851

Qy  1981 GAATGAGGAAGTCTCGCGGAGGCGGCCCAGGAGGCCCTGCGCCCTGAGAACCCGCTCCCGCC 2040
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1852 GAATGAGGAAGTCTCGCGGAGGCGGCCCAGGAGGCCCTGCGCCCTGAGAACCCGCTCCCGCC 1911

Qy  2041 CGAGGAGGGGCGGGGAAGTCAGCCAGGTTGCAGGCCCCGAGCCCGCTCCCGCC 2100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1912 CGAGGAGGGGCGGGGAAGTCAGCCAGGTTGCAGGCCCCGAGCCCGCTCCCGCC 1971

Qy  2101 CTCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGGCCAAGC 2160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1972 CTCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGGCCAAGC 2031

Qy  2161 CTCCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGGCCAAGC 2220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2032 CTCCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGGCCAAGC 2091

Qy  2221 AAAATGCACACCCTTTTTCGTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTTAGAAA 2280
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2092 AAAATGCACACCCTTTTTCGTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTTAGAAA 2151
```

```
Qy  2281 GGGAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCGAGCGGTAATATAT 2340
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2152 GGGAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCGAGCGGTAATATAT 2211

Qy  2341 TTTAGAAATAATATATCAAATAACTAACTAACTCCAATTTTAATCAATTATTAATTATTTT 2400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2212 TTTAGAAATAATATATCAAATAACTAACTAACTCCAATTTTAATCAATTATTAATTATTTT 2271

Qy  2401 TTTTTCTTTTAAAGAGAAGAGCAGGCTTTCTAGACTTAAAGAATAAAGTCTTTGGGAG 2460
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2272 TTTTTCTTTTAAAGAGAAAGCAGGCTTTCTAGACTTAAAGAATAAAGTCTTTGGGAG 2331

Qy  2461 GTCTCACGGTGTAGAGGAGGACTTTGAGCGCCACCCGCACAAAATTCACCCAGAGGGAAAT 2520
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2332 GTCTCACGTGTAGAGGAGGAGCTTTGAGGCCCACCCGCACAAAATTCACCCAGAGGGAAAT 2391

Qy  2521 CTTCGTCGGAAGGACACTCACGCAGTTCTGGATCACCTGTATGTCAACAGAAGAGGGATA 2580
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2392 CTTCGTCCGAAGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAAGAGGGATA 2451

Qy  2581 CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTCATGCCTGTCTAGCTCATACACCCATT 2640
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2452 CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTCATGCCTGTCTAGCTCATCATACACCCATT 2511

Qy  2641 TCTCTTTGCTTCACAGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCT 2700
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2512 TCTCTTTGCTTCACAGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCT 2571

Qy  2701 CTCTCTGTTTATTCTCTCCCCCTCCCCCCTCCCCCCCCCCCCCCCAGGCAAAGCA 2760
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2572 CTCTCTGTTTATTCTCTCCCCCTCCCCCCTCCCCCCCCCCCCCCCAGGCAAAGCA 2631

Qy  2761 TTTCCCATCCTCAATCCCGATATCACACACACAAGGACAAGAGCGAACACACAGCCAGCTCAA 2820
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2632 TTTCCTCATCCCTCAATCCCGATATCACACACACAAGGACAAGAGCGAACACACAGCCAGCTCAA 2691

Qy  2821 GTGCTGAGTATCACATCACAGAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAG 2880
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2692 GTGCTGAGTATCACATCACAGAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAG 2751

Qy  2881 CTTACACTTGGTTACTCAAAGAAAGACCAACCAATCAACCAAAAGAAAAATTCAGCAAAATG 2940
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2752 CTTACACTTGGTTACTCAAAGAAAGACAACCAATCAACCAAAAGAAAAATTCAGCAAAATG 2811

Qy  2941 AGGAAAACAGGAAACCCACCCAAACAACCAATCAACCAAAAGAAAAATTCAGCAAAATG 3000
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2812 AGGAAAACAGGAAACCCACCCAAACAACCAATCAACCAAAAGAAAAATTCAGCAAAATG 2871

Qy  3001 AAAGAATGTATTTGTCTTTTGCATTTGCATTTTGTGTATAATAATTCAGCAAAATG 3060
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2872 AAAGAATGTATTTGTCTTTTGCATTTTGTGTATAATAATTCAGCAAAATG 2931

Qy  3061 ATTCCTTCTTTAAAAAAAAATGTGGAGGAAAGTAGAAAAATTACCAAGGTTGTTGGCC 3120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2932 ATTCCTTCTTTAAAAAAAAATGTGGAGGAAAGTAGAAAATTTACCAAGGTTGTTGGCC 2991

Qy  3121 CAGGGGGTTAAACTTCACCTTGCTCTTGCTCCGATCTCTCCTGAA 3180
         |||||||||||||||||||||||||||||||||||||||||||||
Db  2992 CAGGGGGTTAAACTTCACCTTGCTCTTGCTCCGATCTCTCCTGAA 3051

Qy  3181 GTGTTTTTACCTCAGCACCTTGCTCTTTATTTTTTAATAAAAAATAGAGTAAAGCTGATAG 3240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3052 GTGTTTTTACCTCAGCACCTTGCTCTTTTATTTTTTTAATAAAAAATAAGAGTAAAGCTGATAG 3111

Qy  3241 TTGGAGCATTTTTTTTATTTTTTTAATAAAAAATAAGAGTGGAAAAAATAAGATATCAACT 3300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3112 TTGGAGCATTTTTTTTATTTTTTTAATAAAAAATAAGAGTGGAAAAAAATAAGATATCAACT 3171

Qy  3301 GCCAGCCTGGAGAAGTGAGAAGTGACAGTGCAAGTGTCAAGAGTGTTCGAATTGTCTTCCGCT 3360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3172 GCCAGCCTGGAGAAGTGAGAAGTGACAGTGCAACAGTGTTCAAGAGTGTTCGAATTGTCTTCCGCT 3231
```

Qy   3361 AGCCAAGAACNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3412
            ||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   3232 AGCCAAGAACNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3283

RESULT 8
AF057352
LOCUS   AF057352   3667 bp   mRNA   linear   PRI 23-MAY-1999
DEFINITION Homo sapiens hepatocellular carcinoma autoantigen (p62) mRNA,
            complete cds.
ACCESSION   AF057352
VERSION   AF057352.1 GI:4883680
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3667)
  AUTHORS   Zhang,J.Y., Chan,E.K., Peng,X.X. and Tan,E.M.
  TITLE   A novel cytoplasmic protein with RNA-binding motifs is an
            autoantigen in human hepatocellular carcinoma
  JOURNAL   J. Exp. Med. 189 (7), 1101-1110 (1999)
  MEDLINE   99207072
   PUBMED   10190901
REFERENCE   2 (bases 1 to 3667)
  AUTHORS   Zhang,J.Y., Chan,E.K.L., Peng,X.X. and Tan,E.M.
  TITLE   Direct Submission
  JOURNAL   Submitted (03-APR-1998) Molecular & Experimental Medicine, The
            Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla,
            CA 92037, USA
FEATURES   Location/Qualifiers
     source   1..3667
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     gene   1..3667
                     /gene="p62"
     CDS   436..2106
                     /gene="p62"
                     /note="RNA-binding protein"
                     /codon_start=1
                     /product="hepatocellular carcinoma autoantigen"
                     /protein_id="AAD31596.1"
                     /db_xref="GI:4883681"
                     /translation="MMNKLYIGNLSPAVTADDLRQLFGDRKLPLAGQVLLKSGYAFVD
                     YPDQNWAIRAIETLSGKVELHGKIMEVDYSVSKKLRSRKIQRNIPHLQWEVLDGLL
                     AQYGTEWEQVNTDTETAVVNVTYATREEAKIAMEKLSGHQFENYSFKISYIPDEEV
                     SSPSPPRQRAQRGDHSSREGHAPGSTSQARQIDFPLRILVPTQFVGAIIGKSGLTIKN
                     ITKQTQSRVDIHRKENSGAAEKPVTIHATPEGTSEACRMILEIMQKEADETKLAEEIP
                     LKILANHGLVGRLIGKEGRNLKKIEHETGTKITISSLQDLSIYNPRTITVKGTVEAC
                     ASAEIEIMKKLREAFENDMLAVNTHSGYFSSLYPHHQFGPFBHHSYPEQEIVNLFIP
                     TQAVGAIIGKKGAHIKQLARFAGASIKIAPAEGPDVSERMVIITGPPEAQFKAQGRIF
                     GKLKEENFPNPKEEVKLEAHIRVPSSTAGRVIGKGKTVNELQNLTSAEVIVPRDQTP
                     DENEEVIVRIIGHFFASQTAQRKIREIVQQVKQEQKYPQGVASQRSK"

ORIGIN
Query Match   91.3%; Score 3114.8; DB 9; Length 3667;
Best Local Similarity   96.1%; Pred. No. 0;
Matches 3277; Conservative   0; Mismatches   3; Indels   131; Gaps   3;

Qy   3 CAGCGGCGAGGAGGCGAGGAGGCGAGGGTACCGGCCGCCGGCTCTCGGGGAA   62
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   369 CAACGGAGAGGAGGCGAGGAGGCGAGGGTACCGGCCGCCGGCTCTCGGGGAA 428

Qy   63 GAGACGGATGATGAACAAGCTTACATCGGACAAGCTGAGCCCGTCACCGCCGACGA 122
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   429 GAGACGGATGATGAACAAGCTTTACATCGGACAAGCTGAGCCCGTCACCGCCGACGA 488

Qy   123 CCTCCGGCAGCTCTTTGGGACAGGAAGCTTTGGGGACAGGAAGCTGCCCCTG... 182
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   489 CCTCCGGCAGCTCTTTGGGACAGGAAGCTTTGGGGACAGGAAGCTGCCCCTG... 548

Qy   183 CGGCTACGGCCGCCTTCGTGACTACCCCGACCAGAACTGGGCCACTCCGGCATCGAGACCCT 242

Qy   549 GGCTACGCCTTCGTGGACTACCCCGACCAGAACTGGGCCATCCGCGCCATCGAGACCCT 608
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   243 CTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTGATTACTCAGTCTCTAAAAA 302

Db   609 CTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTGATTACTCAGTCTCTAAAAA 668

Qy   303 GCTAAGGAGCAGAAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGGAGGTGTT 362
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   669 GCTAAGGAGCAGAAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGGAGGTGTT 728

Qy   363 GGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGGAACAGTCAACACAGACAC 422
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   729 GGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGGAACAGTCAACACAGACAC 788

Qy   423 AGAAACCGCCGTTGTCAACGTCACATATGCAACAGAGAGAGCAAAAATAGCCATGGA 482
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   789 AGAAACCGCCGTTGTCAACGTCACATATGCAACAGAGAGAGCAAAAATAGCCATGGA 848

Qy   483 GAAGCTAAGGCGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGATGA 542
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   849 GAAGCTAAGGCGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGATGA 908

Qy   543 AGAGGTGAGCTCCCTTGCCCCCTCAGCGAGCCCAGGGTGGGGACCACTCTTCCCGGGA 602
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   909 AGAGGTGAGCTCCCTTGCCCCCTCAGCGAGCCCAGGGTGGGGACCACTCTTCCCGGGA 968

Qy   603 GCAAGGCCACGCCCCTGGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGGAT 662
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   969 GCAAGGCCACGCCCCTGGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGGAT 1028

Qy   663 CCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAA 722
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1029 CCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAA 1088

Qy   723 CATCACTAAGCAGACCCAGTCCCGGTAGATATCCATAGAATCTCGAGCTGC 782
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1089 CATCACTAAGCAGACCCAGTCCCGGTAGATATCCATAGAATCTCGAGCTGC 1148

Qy   783 AGAGAAGCCTGTCACCATCCATGCACCCCAGAGGGGACTTCTGAAGCATGCGCGCATGAT 842
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1149 AGAGAAGCCTGTCACCATCCATGCACCCCAGAGGGGACTTCTGAAGCATGCGCGCATGAT 1208

Qy   843 TCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAA 902
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1209 TCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAA 1268

Qy   903 AATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAGAAGGGCAGAAATTTGAA 962
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1269 AATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAGAAGGGCAGAAATTTGAA 1328

Qy   963 GAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTGAGCAT 1022
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1329 GAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTGAGCAT 1388

Qy   1023 ATACAACCCGGAAAGAACCACACTGTGAAGGGCACAGTGAAGGCCTGAGCCTGTCCAGTGCTGA 1082
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1389 ATACAACCCGGAAAGAACCACCATCACTGTGAAGGGCACAGTGAAGGCCTGAGCCTGTCCAGTGCTGA 1448

Qy   1083 GATAGAGATTATGAAGAAGCTGCGTGAGGCCTTGAAAATGATATGCTGGCTGTTAACCA 1142
            |||||||||||||||||||||||||||||||||||| |||||||||||
Db   1449 GATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTGTTAAC-- 1506

Qy   1143 ACAAGCCAATCCAATCTGATCCCCAGAGGTTGAACCTCAGCGCACTTGGACATCTTTTCAACAGGACT 1202
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1507 --- 1507

Qy   1203 GTCCGTGCTATCTCCACCAGACAGGGACCCCGCGGGCCCCGGAGCTCCCCCGTCACCACCC 1262
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1507 --- 1507

Qy   1263 CTTCACTACCCCACTCCGATACTTCTCCAGCTGTACCCCATCACCAGTTGCCTGAAGTC 1322
            ||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db 1507 --------ACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGCCCGTT 1559
Qy 1323 CCGGCATCATCACTCACTCTTATCCAGACGAGGAGATTGTGAATCTCTTCATCCAACCAGC 1382
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1560 CCCCGGCATCATCACTCACTCTTATCCAGACGCAGGAGATTGTGAATCTCTTCATCCAACCCAGC 1619

Qy 1383 TGTGGGCGCCATCATCGGGAAGAGAGGGGCACACATCAAACAGCTGGCGAGATTCGCCGG 1442
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1620 TGTGGGCGCCATCATCGGGAAGAGAGGGGCACACATCAAACAGCTGGCGAGATTCGCCGG 1679

Qy 1443 AGCCTCTATCAAGATTGCCCCTGCCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCAT 1502
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1680 AGCCTCTATCAAGATTGCCCCTGCCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCAT 1739

Qy 1503 CACCGGGCACCCGAAGCCCAGTTCAAGGCCCAGGGACCGGATCTTTGGGAAACTGAAAGA 1562
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1740 CACCCGGGCACCCGAAGCCCAGTTCAAGGCCCAGGGACCGGATCTTTGGGAAACTGAAAGA 1799

Qy 1563 GGAAAACTTCTTTAACCCCAAAGAGAGAAGTGAAGCGCATATCAGAGTGCCCTC 1622
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1800 GGAAAACTTCTTTAACCCCAAAGAGAGAAGTGAAGCGCATATCAGAGTGCCCTC 1859

Qy 1623 TTCCACAGCTGGCCGGTGATTGGCAAAGTGGCAAGACCGTGAACGAACTGCAGAACTT 1682
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1860 TTCCACAGCTGGCCGGTGATTGGCAAAGTGGCAAGACCGTGAACGAACTGCAGAACTT 1919

Qy 1683 AACCAGTGCAGAAGTCATCGTGCCCTCGTGACCAAACGCAGATGAAAATGAGGAAGTGAT 1742
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1920 AACCAGTGCAGAAGTCATCGTGCCCTCGTGACCAAATCGCAGATGAAAATGAGGAAGTGAT 1979

Qy 1743 CGTCAGAATATCGGCCACTCTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAAT 1802
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1980 CGTCAGAATTATCGGCCACTCTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAAT 2039

Qy 1803 TGTACAACAGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCCAG 1862
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2040 TGTACAACAGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCCGCAG 2099

Qy 1863 CAAGTGAGGCTCCACAGGCACCAGCAGCACCGGATGAATGTAGCCCTTCCAACACCT 1922
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2100 CAAGTGAGGCTCCACAGGCACCAGCAGCACCGGATGAATGTAGCCCTTCCAACACCT 2159

Qy 1923 GACAGAATGAGACCAAAACGCAGCCAGATCGGGAGCAAACCAAGACCATCTGAGGA 1982
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2160 GACAGAATGAGACCAAAACGCAGCCAGATCGGGAGCAAACCAAGACCATCTGAGGA 2219

Qy 1983 ATGAGGAAGTCTGCGGAGGGGCGGCCAGGGGCCAGGACTCCTGCCGAGACCCCAGGGGCCG 2042
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2220 ATGAGGAAGTCTGCGGAGGGGCGGCCAGGGGCCGCCGAGACCTCGCCGAGACCCCAGGGGCCG 2279

Qy 2043 AGGAGGGGGGGGGGAAGGTCAGCAGGTTGCCAGAAGGCCCCGCTCCTCCGCCCC 2102
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2280 AGGAGGGGGGGGGGAAGGTCAGCAGGTTGCCAGAACCACCGAGGCCCCGCTCCTCCGCCCC 2339

Qy 2103 CCCACGACGCTATCCCTTTTGTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGACT 2162
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2340 CCCACGACGCTATCCCTTTTGTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGACT 2399

Qy 2163 AATGCACACCCCTTTTCTGTGGCAAATGTCTCTGTACATGTGTACATATTAGAGAAGG 2222
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2400 AATGCACACCCCTTTTCTGTGGCAAATGTCTCTGTACATGTGTACATATTAGAGAAGG 2459

Qy 2223 GAAGATGTTAAGATAGTATGTGCCCTGGGTTACACAGGGTGCCTGCAGCGGTAATATATTT 2282
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2460 GAAGATGTTAAGATAGTATGTGCCCTGGGTTACACAGGGTGCCTGCAGCGGTAATATATTT 2519

Qy 2343 TAGAAATAATAATATCAAATAACTCAACTAACTCCAATTTTAATCAATTAATTATTAATTTTTT 2402
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2580 TAGAAATATATATAATATCAAATAACTCAACTAACTCCAATTTTAATCAATTAATTATTAATTTTTT 2639
```

```
Qy 2403 TTTCTTTTTAAAGAGAAAGCAGCGTTTTCTAGACTTAAAGAATAAAGTCTTTGGGAGGT 2462
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2640 TTTCTTTTTAAAGAGAAAGCAGCGTTTTCTAGACTTAAAGAATAAAGTCTTTGGGAGGT 2699

Qy 2463 CTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGACACAAAATTCACCCAGAGGGAAATCT 2522
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2700 CTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGACACAAAATTCACCCAGAGGGAAATCT 2759

Qy 2523 CGTCGGAAGGACACTCACGGCAGTTCTTGGATCACCTGTGATGTCAACAGAAGGGATACC 2582
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2760 CGTCGGAAGGACACTCACGGCAGTTCTTGGATCACCTGTGATGTCAACAGAGGGGATACC 2819

Qy 2583 GTCTCCTGAAGAGGAAACTCGTCACTCCTCATGCCTGTCGTCAGCTCATCACCCATTTC 2642
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2820 GTCTCCTGAAGAGGAAACTCGTCACTCCTCATGCCTGTCGTCAGCTCATCACACCCATTTC 2879

Qy 2643 TCTTTGCTTCACAGGTTTAAACTGGTTTTTTGCATACTGCTATATACTTCTGTCTCT 2702
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2880 TCTTTGCTTCACAGGTTTAAACTGGTTTTTTGCATACTGCTATATATTTCTGTCTCT 2939

Qy 2703 CTCTGTTTATCTCTCTCCCCCTCCCCTTCTTCTCCATCTCCATTCTTTTTTGAATT 2762
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2940 CTCTGTTTATCTCTCTCCCCTCCCCTTCTTCTCCATCTCCATTCTTTTTTGAATT 2999

Qy 2763 TCCTCATCCCTCCATCTCAATCCCGTATCTCACGCACCCCCCCCCCCCCAGGCAAAGCAGT 2822
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3000 TCCTCATCCCTCCATCTCAATCCCGTATCTCACGCACCGA-CCCCCCCCCCAGGCAAAGCAGT 3058

Qy 2823 GGTCTGAGTATCACATCACACAAAAAGGGAAACACACAAAACCAGCCTCAACT 2882
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3059 GGTCTGAGTATCACATCACACAAAAAGGGAAACACACAAAACCAGCCTCAACT 3118

Qy 2883 TACACTTGGTTACTCAAAAGAACAAGACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAGAG 2942
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3119 TACACTTGGTTACTCAAAAGAACAACACAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAGAG 3178

Qy 2943 GAAAACAGGAACCCACCCAAACCAATCAACCAACAAGAAAAATTCCACAATGAA 3002
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3179 GAAAACAGGAACCCACCCAAACCAATCAACCAACAAGAAAAATTCCACAATGAA 3238

Qy 3003 AGAATGTATTTTGTCTTTTGGTGTATAAGCCATCAATAATTCAGCAAAATGAT 3062
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3239 AGAATGTATTTTGTCTTTTGGTGTATAAGCCATCAATATTCAGCAAAATGAT 3298

Qy 3063 TCCTTTCTTT-AAAAAAAAAATGTGAGGAAATTACCAAGGTTGTTGCCCC 3121
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3299 TCCTTTCTTTAAAAAAAAAATGTGAGGAAATTTACCAAGGTTGTTGCCCC 3358

Qy 3122 AGGGCGTTAAATTCACAGATTTTTTTAACGAGAAAAACACACAGAAGAAGCTACCTCAGG 3181
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3359 AGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAAGAAGCTACCTCAGG 3418

Qy 3182 TGTTTTTACCTTGCTCTCTGTTCCCTTAGAGATTTTGGAAAAATAAGATATCAACTG 3241
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3419 TGTTTTTACCTTGCTCTCTGTTCCCTTAGAGATTTTGGAAAAATAAGATATCAACTG 3478

Qy 3242 CCAGCCTGGAGAAGGTGACAGTCCAAGTGTCGCAAGAGTTCTCGCTA 3301
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3479 CCAGCCTGGAGAAGGTGACAGTCCAAGTGTCGCAAGAGTTCTCGCTA 3538

Qy 3302 CCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATT 3361
       |||||||||||||||||||||||||||||||||||||||||||||||||
Db 3539 CCAAGAACCTATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATT 3598
```

```
Qy 3362 GCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATT 3412
       ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3599 GCCAAGAACCTATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATT 3649
```

RESULT 9
AC020629
LOCUS   AC020629   142971 bp   DNA   linear   PRI 07-MAR-2002

```
DEFINITION   Homo sapiens 12q BAC RP11-76E16 (Roswell Park Cancer Institute
             Human BAC library) complete sequence.
ACCESSION    AC020629
VERSION      AC020629.6   GI:7656675
KEYWORDS     HTG.
SOURCE       Homo sapiens (human).
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 142971)
AUTHORS      Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
             Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
             Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
             David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
             Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
             Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
             Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
             Holloway,C., Hosak,H., Issar,A., Jackson,L.E., Jackson,L., Jia,Y.,
             Jones,M., Kelly,S., Kneitz,S., Kondejewski,N., Kong,Y., Kovar,C.,
             Lau,S., Leal,B., Lee,E., Li,J., Lichtarge,O., Liu,J., Liu,W.,
             Logan,O., Lozado,R.J., Lu,J., Lucier,R., Marondel,I., Martin,R.,
             Martinez,C., McLeod,M.P., Mei,G., Merscher,S., Miller,A.,
             Montgomery,K.T., Morgan,M., Morris,S., Nash,S., Nelson,A.,
             Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B., Paxton,S.,
             Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,D., Rives,M.,
             Samuel,S., Say,C., Scherer,S., Shah,F., Shen,H., Shim,C., Simon,M.,
             Sparks,A., Stamps,A., Sucgang,R., Tabor,P., Taylor,T., Vasquez,L.,
             Vinson,R., Vo,Q., Wahbah,M., Watlington,S., Weinstock,G.,
             Weinstock,I.R., Williamson,A., Worley,K., Wren,J., Wrensford,G.,
             Xhang,A.M., Yang,R., Yu,W., Zhou,X., Kucherlapati,R., Nelson,D. and
             Gibbs,R.
TITLE        Direct Submission
JOURNAL      Unpublished
REFERENCE    2  (bases 1 to 142971)
AUTHORS      Worley,K.C.
TITLE        Direct Submission
JOURNAL      Submitted (07-JAN-2000) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
REFERENCE    3  (bases 1 to 142971)
AUTHORS      Worley,K.C.
TITLE        Direct Submission
JOURNAL      Submitted (27-APR-2000) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
REFERENCE    4  (bases 1 to 142971)
AUTHORS      Worley,K.C.
TITLE        Direct Submission
JOURNAL      Submitted (28-APR-2000) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
REFERENCE    5  (bases 1 to 142971)
AUTHORS      Worley,K.C.
TITLE        Direct Submission
JOURNAL      Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
COMMENT      On Apr 27, 2000 this sequence version replaced gi:7025656.
             INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
             gc-help@bcm.tmc.edu

             CLONE LENGTH: This sequence does not necessarily represent the
             entire insert of this clone. Overlapping regions of clones are only
             sequenced and submitted once, so the sequence for the remainder of
             the insert may be found in the record for the adjacent clones.
             Overlapping clones are noted at the beginning and end of the
             Features listing.

             ANNOTATION OF FEATURES:
             STSs are identified using ePCR (Genome Res. 7:541-550) searches
             of a local database that includes entries from dbSTS, GDB, and
             local mapping efforts.
             Repeats are identified using RepeatMasker (A. Smit and P. Green,
```

```
             unpublished.) for Human and Mouse sequences.
             Genes and Region of sequence similarity are identified by BLAST
             (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
             EST and cDNA sequences. Genes demonstrate at least two exons
             flanked by consensus splice sites that maintained sequence
             continuity across the splice junctions. Sequences that are not
             identical matches are annotated as similar.

             SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
             standard of double strand coverage with a minimum of 2 clones and 2
             reads with no ambiguities or 2 chemistries with a minimum of 2
             clones and 3 reads with no ambiguities. If the sequence quality for
             a region does not meet this standard, it will be indicated in the
             annotation as Low Coverage.

             QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
             standards - estimated error rate less than 1 per 10,000 bases.
             Reports of lowest quality individual bases and measures of base
             quality are listed below. Description of the metrics can be found
             at URL:
             http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

             QUALSTAT-REPORT-------

             ------------- Summary Statistics -------------
             Contig length:                               142971
             Phrap values in estimate:                    140680
             Average error rate:                          0.000118703
             Average error rate (BCM-Phrap estimate):     0.0433608
             Fraction of Phrap values less than 40 :      31
             Number of consensus changing edits:          0
             Number of N's in consensus :

             ------------- Consensus changing edits -------------
Position     Original+Context           Edited+Context
11171        gtcttctgga(n)caaaaaaaa      gtcttctgga(a)caaaaaaaa
12018        tactcatctg(n)ctatatctat     tatccatctg(t)ctatatctat
16604        tccactctaa(n)agttccatt      tccactctaa(g)agttccatt
18385        tagtaaccac(n)tacattttta     tagtaaccac(t)tacattttta
21879        aaaaaaaaaa(n)nttaaaaca      aaaaaaaaaa(a)attaaaaca
21880        aaaaaaaaaa(n)nttaaaaca      aaaaaaaaaa(a)attaaaaca
21881        aaaaaaaaann(n)ttaaaaacag    aaaaaaaaann(a)ttaaaaacag
21906        cctaggtccc(n)cttaaatgag     cctaggtccc(a)cttaaatgag
22169        accagcctgg(n)caacatggcg     accagcctgg(c)caacatggcg
38915        ttacatctgt(n)tctttgcaaa     ttacatctgt(t)tctttgcaaa
55675        acttgtctc(n)aaaaaaaaaa       acttgtctc(a)aaaaaaaaaa
57658        gaccagcctg(n)ccaacatggt     gaccagcctg(a)ccaacatggt
74371        gtgtcggccc(n)cgcccgcgc      gtgtcggccc(c)cgcccgcgc
88021        taaatgaggg(n)agaccagtgg     taaatgaggg(c)agaccagtgg
88248        gcagtgagcc(n)agatcacacc     gcagtgagcc(a)agatcacacc
94121        ttgttcccga(n)tgctctctgc     ttgttcccga(t)tgctctctgc
95228        gtccatgtgt(n)cnnttgttc      gtccatgtgt(t)ctcattgttc
95230        ccatgtgtnc(n)nnttgttcaa     ccatgtgtnc(t)cnttgttcaa
95231        catgtgtncn(n)nttgttcaac     catgtgtncn(n)ttgttcaac
95232        atgtgtncnn(n)ttgttcaact     atgtgtncnn(n)ttgttcaact
104931       tttttttag(n)acagacagg       tttttttag(t)acagacagg
118168       gcccaggtcc(n)ccaaggtgtt     gcccaggtcc(n)ccaaggtgtt
120974       catctcaaaa(n)ananangaac     catctcaaaa(a)aaaaaagaac
120976       tctcaaaana(n)anangaactc     tctcaaaana(a)anangaactc
120978       tcaaaanana(n)angaactctt     tcaaaanana(a)angaactctt
120980       aaaananana(n)gaactcttcc     aaaaaaaaaa(a)gaactcttcc
139505       ccagccatgg(n)gggggcacc      ccagccatgg(t)gggggcacc
139600       gccatgggg(n)gggcacctgt      gccatggg(t)gggcacctgt
139698       attaccatta(n)aatatggtat     attaccatta(g)aatatggtat
140104       accgtgccca(n)cagaatagaa     accgtgccca(g)cagaatagaa
140149       tggaacatt(n)gataccaca       tggaacatt(g)gataccaca

             ------- Distribution of Quality < 40 Bases -------
                 1000|                                    *    *
                  900|                                    *    *
```

```
          800
  #       700
  bases   600
          500
          400
          300
          200
          100
            0
              |----|----|----|----|----|----|----|----|
              5   10   15   20   25   30   35   40
                        Phrap Value Range
```

```
FEATURES        Version: 1.01 qxfo.
                Location/Qualifiers
source          1..142971
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="12"
                /clone="RP11-76F16"
repeat_region   147..236
                /rpt_family="A-rich"
repeat_region   258..378
                /rpt_family="AluJb"
repeat_region   complement(537..840)
                /rpt_family="AluSx"
repeat_region   complement(1722..2137)
                /rpt_family="Tigger2a"
repeat_region   complement(3161..3584)
                /rpt_family="MSTD"
STS             3228..3502
                /standard_name="G05956"
                /db_xref="dbSTS:11453"
STS             3347..3490
                /standard_name="cda0yb04"
                /db_xref="GDB:446196"
misc_feature    3585..3735
                /note="Region similar to Hs#S936143 aj27d02.s1 Homo
                sapiens cDNA: AA789222"
repeat_region   3885..3910
                /rpt_family="(TA)n"
repeat_region   complement(3917..4030)
                /rpt_family="AluB"
repeat_region   complement(4544..4681)
                /rpt_family="MIR"
misc_feature    5145..5510
                /note="Region similar to Hs#S953417 oa06g05.s1 Homo
                sapiens cDNA: AA765439"
misc_feature    5553..5859
                /note="Region similar to Hs#S1195593 qt69f10.x1 Homo
```

```
Query Match       78.0%;   Score 2662;   DB 9;   Length 142971;
Best Local Similarity   90.2%;   Pred. No. 0;
Matches 3085;   Conservative   0;   Mismatches 156;   Indels 180;   Gaps 15;
```

```
Qy       1  GGCAGCGGAGGAGGCGAGAGCGCCGGGTACCGGCCGCCGGGGAGCCGCGGGGCTCTCGGGG   60
            ||||||  |||||||||||||| ||||||  ||||||  |||||||  |||  ||||||||
Db  132954  GGCAGCAGAGGAGGAGCGAGAGGCGCAGGGGTATCGGG-CGGGGGAGCCCCTCGGGG  133012

Qy      61  AAGAGACGGATGATGAACAAGCTTTACATCGGGAGGAAGCTGAGCCCCGTCACCGCCGAC  120
            ||| ||| ||||||||||||||||||||  |||||  ||||| | | ||||||||||
Db  133013  AACAGACGGATGATGAACAAGCTTTTCATCGGGAGGAAGCTGCCCCCGTCACCGCGCCGAA  133072

Qy     121  GACCTCCGCAGCTCTTTGGGGACAGGAGAAGCTGCCCGCGGGGACAGGTCCTGCTGAAG  180
            ||||||||||||||||||||||||| ||||||||||| | | | |  |||||||||||||
Db  133073  GACCTCCGCAGCTCTTTGGGGACTACCCCGACCAGGAAGCTGCCCGCGGGGACAGGTCCTGCTGAAG  133132

Qy     181  TCCGGCTACGCCTTCGTGGACTACCCCGACTACCCCAGCTTCTCCAGCCTGTACCCCCATCACCAGTTTGGCCCG  240
            ||  |||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||
Db  133133  TCCCGCTACGCCTTCGTGGACTACCCCGACTACCCCAGCTTCTCCAGCCTGTACCCCCATCACCAGTTTGGCCCG  133192
```

```
Qy     241  CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133193  CTCTCGGGTCAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTATTCAGTCTCTATA  133252

Qy     301  AAGCTAAGGAGCAGGAGAAATCAGATTCGAAACATCCCTCGTCACCTGCAGTGGGAGGGTG  360
            ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db  133253  AAGCTAAGGAGCAGGAGAACATTCGATTCGAAATATCCCTCGTCACCTGCAGTGGGAGGGTG  133312

Qy     361  TTGGATGGACTTTGGCTCAATATGGACAGTGGAACAGTGGAACAAGTCAACACAGAC  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133313  TTGGATGGACTTTGGCTCAATATGGACAGTGGAACGTGGAACAAGTCAACACAGAC  133372

Qy     421  ACAGAACCGCGTTGTCAAGTCACATATGCAACATATGCAACAAAAGAGAAGAGAAGTAAAAATAGCCATG  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133373  ACAGAACACGCTGTTGTCAAGTCACATATGCAACATATGCAACAAAAGAGAAGAGAAGTAAAAATAGCCATG  133432

Qy     481  GAGAAGCTAAGGCGCATCAGTTGAGAACTACTCCTTCAGATTCCTACATCCCGGAT  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133433  AAGAAGCTAAGGCGCATCAGTTTGAGAACCACTACTTCAGATTTCCTACATCCCGGAT  133492

Qy     541  GAAGAGGTGAGCTCCCCTTCGCCCCCTCAGCGAGCCCCAGCGTGGGGACCACTCTTCCGG  600
            || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133493  GACGAGGTGAGCTGCCCTTCGCCCCCTCAGCGAGCCCCAGCGTGGGGACCACTCTTCCTGG  133552

Qy     601  GAGCAAGGCCACGCCCTGGGGGCACTTCTCAGCGCCAGACAGATTCCCGTGCGG  660
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133553  GAGCAAGGCCAGCCCTGGGGGCTCTTCTCAGCGCCAGACAGATTGATTTCCCACTGCGT  133612

Qy     661  ATCCTGGTCCCCACCAGTTTGTTGGCCATCATCGGAAAGGAGGGCTTGACCATAAAG  720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133613  GTCCTGTTCCCCACCAGTTTGTTGGCCATCATCGGAAAGGAGGGCTTGACCATAAAG  133672

Qy     721  AACATCACTAAGCAGACACCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT  780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133673  AACATCACTAAGCAGACCCGGGTCCCGGGTAGACATCTATAGACAAGAGAACTCCAGAGCT  133732

Qy     781  GCAGAGAAGCCTGTCACCATCCACCCCAGAGGGGACTTCTGAAGCATGCGCCATG  840
            |||||||||||||||||           ||||||||||||||||||||||||||||
Db  133733  GCAGAGAAGCCTGTCA----CCATGCCACCCCAGAGGGGACTTCTGAAGCATGCGCCATG  133788

Qy     841  ATTCTTGAAATCATGCAGAAAGAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTG  900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133789  ATTCTTGAAATAATGCAGAAAGAGAGGCAGATGAGCCGCAAACTAGCCGAAGAGATTCCTCTG  133848

Qy     901  AAAATCTTGGCACAACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTG  960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133849  AAAATCTTGGCCCACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTG  133908

Qy     961  AAGAAATTGAACATGAAACAGGGACCCAAGATAACAATCTCATCTTCGCAGGATTTGAGC  1020
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133909  AAGAAAAATGAACATGAAACAGGGACCCAAGATAACAATCTCATCTTCGCAGGATTTGAGC  133968

Qy    1021  ATATACAACCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTTGCCAGTGCT  1080
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133969  ATATACAACCGGAAAGAACCATCACTGTGAAGGGCACAGTCGTGCCAGTGCT  134028

Qy    1081  GAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGTCTGGTTAAC  1140
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  134029  GAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGTCGACTGTTAA·  134087

Qy    1141  CAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGACTTCTTTTCAACAGGA  1200
Db  134088  -----------------------------------------------  134087

Qy    1201  CTGTCCGTGCTATCTCCACCAGCAGGCTGCTGAAG  1260
            |||||||||||||||||||||||||||||||||||
Db  134088  CTGTCCGTGCTATCTCCACCAGCAGGCTCCTGCTGAAG  134087

Qy    1261  CCCTTCCTACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTTGGCCCG  1320
            |||||||||||||||||||||||||||||||||||||||
Db  134088  --------TACCCACTTCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTTGGCCCG  134139
```

```
Qy      1321 TTCCCGCATCATCACTCTTATCCAGAGCAGAGAGATTGTGAATCTTCATCCCAACCCAG 1380
             ||||||||||||| ||||||||||||||||||||||||||||| |||||||||||||||
Db    134140 TTCCCGCATCATCACTCTTATCCAGAGCAGAGAGATTGTCATCATTCATCCCAACCCAG 134199

Qy      1381 GCTGTGGGCGCCATCATCGGGAAGAAGGGGCACACATCAAACAGCTGGCGAGATTCGCC 1440
             |||||||||||||||||||||||||||||| |||||||||||||||||||||||||| |
Db    134200 GGTGTGGGCGCCATCATCGGGAAGAAAGGGGCACACATCAAACAGCTGGCGAGATTCGTG 134259

Qy      1441 GGAGCCTCTATCAAGAGATTGCCCCTGCGGAAGGCCCAGAGCGTCAGCGGAAGGATGGTCATC 1500
             ||||||||||||||||||| |||||||                  CGGAAGGTCATC
Db    134260 GGAGCCTCCATCAAGAGATCGCCCCTGCGGTCAG---------CGGAAGGTCATC 134302

Qy      1501 ATCACCGGGCACCGGAAGCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAA 1560
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    134303 ATCACCTGGCACCGGAATCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAA 134362

Qy      1561 GAGGAAAACTTCTTTAACCCAAAGAAGAAGACGTGAAGCGTGGAAGCGCATATCAGAGTGCCC 1620
             |||||||||||||||||||||||||||||||||||||||||||||||||
Db    134363 GAAGAAAACTTTTTTAACCCAAAGAAGAAGACGTGAAGCGTGGAAGCGCATATCAGAGTGCCC 134422

Qy      1621 TCTTCCACAGCTGGCCGGGTGATTGGCAAGAAGGTGCAAGACCGTGAACGAACTGCAGAAAC 1680
             |||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    134423 TCTTCCACCGCTGGCCGGGTGATTGGCAAGAAGGGGCAAGACCGTGAATGAACTGCAGAAT 134482

Qy      1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTCGTGAGCCAGAAACCCAGATGAAAATGAGGAAGTG 1740
             |||||||||||||||||| ||||||||||||||||||||||||||||||
Db    134483 TTAATCAGTGCAGAAGTCATCGTGCCTCGTCGTGAGCCAGAAACGCCAGATGAAAATGAGGAAATG 134542

Qy      1741 ATCGTCAGAATTATCGGGACTTCTTTGCTAGCCAGAAATCACAGCCGCAAGATCAGGGAA 1800
             ||||||||||||||||||||||||||||||||| ||||||||||||
Db    134543 ATCGTCAGAATTATCGGGACTTCTTTGCTAGCCAGAAATACCCAGCCGCAAGATCAGGGAA 134602

Qy      1801 ATTGTACACAGGTGAAGCAGCAGGAGCAGAGAAATACCCTCAGGGAGTCGCCTCACAGCGC 1860
             ||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db    134603 ATTGTACACAGGTGAAGCAGCAGGAGCAGAGAAATACCCTCAGGGAGTCGCCTCACAGCGC 134662

Qy      1861 AGCAAGTGAGGCTCCCACAGGGCAACGAACCCAGCAAACACACGGATGAATGTAGCCCTTCCAACAC 1920
             |||||||||||||||||||||||||||||||||||
Db    134663 AGCAAGTGAGGATCTGCGGGGGAAGCCAACCCAGCAAACACAACGGATGAATGTAGCCCTTCCAACAC 134722

Qy      1921 CTGACAGAATGAGACACAAACGCAGCCAGATCGGGACAACCAAGACCAATCTGAG 1980
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    134723 CTGACAGAATGAGACACAACAGCGCAGCCAGATCAAGAGCAACCAAGACCAATCTGAG 134782

Qy      1981 GAATGAGAAGTCTCGGCGGAGGCGGGAAGTTGCCGAGGACTCTGCCGAGGAACCCGAGGGGC 2040
             ||||||||||||||||||||||||||||||||||||||||||||||
Db    134783 GAATGAGAAGTCTCGCGCGGGGGGCGTTAAACCAGGTTGCCGAGGACTCTGCCGAGGAACCCGAGGGGC 134842

Qy      2041 CGAGGAGGGGCGCGGGGAAGTTAAGGAGGTAAACAGGTTGCCCGAGAACCACCGGGAACACCCGGCCCCC 2100
             ||||||||||||||||||||||||||||||||||||||
Db    134843 CGAGGAGGGGCGCGGGGAAGTTGGGGAAGTAAACAGGTTGCCCGAGAACCACCGGGGAACACCCGGCCCCC 134902

Qy      2101 CCCCAGGGGCTTCTGCAGCCTTCAGCCATCCACTCCACTCGGGATCTCCTCCTTAA 2160
             ||||||||||||||||||||||||||||||||| |||||||||||||
Db    134903 CCCCAGGGGCTTCTGCAGCCTTTGTGGCAAATAGTCCTCCACTCACCATCCTCGGATCTCCTCCTTAA 134962

Qy      2161 CTCCCACGACGCTATCCCTTTGAGTTGAACATAGGTGAAC-GTGTTCAAAGCCAAG 2219
             ||||||||||||||||||||||||||||| || ||||||| |||||||||||||||
Db    134963 CTCCCATGACGCTATCCCTTTAGTTGAACATAGGTAAACATTGTTCAAAGCCAAG 135022

Qy      2220 CAAAATGCACACCCTTTTCTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATTATTAGAA 2279
             |||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    135023 CAAAATCACACGCCCCTTT-TGTGGCAAATAGTCTTCCACACGTGTGTACATTATTAGAA 135081

Qy      2280 AGGGAAGATGTAAGACTATGTGGCCTGTCGGGCCTTGCCTGCCGACAGGGTGCCTGCAGCCGGTAATATA 2339
             ||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db    135082 GGGGAAGATGTAAGACTATGTGGCCTGTGGGGCCTTGCCTGCCGACAGGGTGCCTGCAGCCGGTAATATA 135141

Qy      2340 TTTTAGAAATAATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAAATTT 2399
             |||||||||||||||||||||    -AACTAACTCCAATTTTAATCAATTATTA---T
Db    135142 TTTTAGAAATATATATCAAAT-----AACTAACTCCAATTTTAATCAATTATTA---T 135193

Qy      2400 TTTTTTTTCTTTTAAAGAGAAAGCAGGCTTTTCGACTTTTAAAGAATAAAGTCTTTGGGA 2459
```

```
Qy      2460 TTTTTTCTTTTAAAGAGAGAGGAGGAGGCTTTGTGAGGCACGGCACCCGACACAAAATTCACCCAGAGGAGAAA 2519
             ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db    135194 TTTTTTCTTTTAAAGAGAGAGGAGGAGGCTTTGTGAGGCACGGCACCCGACACAAAATTCACCCAGAGGAGAAA 135250

Qy      2520 GGTCTCACGTGTAGAGGACACTCACGGCAGTTCTGGATCACCGTGTATGTCAACGAAGGAT 2579
             ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    135251 GGTCTCACGTGTAGAGGACACTCACGGCAGTTCTGGATCACCGTGTATGTCAACGAAGGAT 135310

Qy      2580 TCTCGTGGAAGGACACTCACGGCAGTTCTGGATCACCGTGTATGTCAACGACAGAAGGGAT 2639
             |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db    135311 TTTTGTCGGAAGGACACTCACGGCAGTTCTGGATCATGTATGTCAACGACAGAAGGGAT 135370

Qy      2640 ACCGTCTCCTTGAAGAGAAACTCTGTCACTCGTCACTCCTGCTCGTCTAGCTCATCACACCCAT 2699
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    135371 ACTGTCTCCTTGAAGGGGGAAACTCTGTCACTCGTCACTCCTGCTCGTCTAGGTCATCACACCCAT 135430

Qy      2700 TTCTTTGCTTCACAGGTTTAAACTGGTTTTGCATACTGCTATATAATTCTCTGTC 2759
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    135431 TTCTTTGCTTCACAGGTTTAAACTGGTTTTGCATACTGCTATATAATTCTCTGTC 135490

Qy      2760 TCTCTCTGTTTATCTCTCCCCTCTTCTTCTCTCATCTCCATTCTTTTGA 2811
             |||||| --TATCTCTCCCCTCTTCTTCTCTCATCTCCATTCTTTTGA
Db    135491 TCTCTC----TATCTCTCCCCTCTTCTTCTCTCATCTCCATTCTTTTGA 135546

Qy      2812 ATTTCCTCATCCCTCCAATCCGTATCTACGCAC------CCCCCCCCCA 2871
             |||||||||||||||||||||||||||||||| |||||||| ||
Db    135547 ATTCCCTCATCCCTCCATCTCAATCCCATATCTACGCACCACCCCCCTCTGCCA 135606

Qy      2872 GGCAAAGCAGTGCTCTGAGTATCACATCACACAGAACAACAAAAGCGAAACACACAAAC 2931
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    135607 GGCAAAGCAGTGCTCTGAGTATCACATCACACAGAACAACAAAGCGAAACACACAAAC 135666

Qy      2932 CAGCCTCAACTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCG 2991
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    135667 CAGCCTCAACTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCA 135726

Qy      2992 TTTTGGAAGAGGAAAAACAGGAACCCCTTCCACACAC 3051
             ||||||||||||||||||||||||||||||||||
Db    135727 TTTTGGAAGAGGAAAAACAGGAACCCCTTCCACACAC 135786

Qy      3052 TCCACAATGAAAGAAATGTATTTGTCTTTTGTCATTTGGTGTATAAGCCATCAATATTC 3110
             |||| |||||||||||||| ||||||||||||||||||||||||||||||||||
Db    135787 TCCACAATGAAAGAAATGTATTTCTTTTTT-AAAAAAAAATGTGGTGTATAAGGCCATCAATACTC 135846

Qy      3111 AGCAAAATGATTCCTTCTTAAAAAAAAATGTGGAGGAAGTAGAAATTTA-CCAAG 3169
             ||||||||||||||||||||||||||||||||| |||||||||||||| |||||
Db    135847 AGCAAAATAATTTCTTTCTTT-AAAAAAAATGTGGAGGAAGTAGAAATTTATACAAG 135905

Qy      3170 GTTGTTGGCCCAGGCCGTTAACGACGACCAGCACTTCCCTTGTCTTTGTCTTTCCTGTGTC 3228
             ||||||||||||||| |||||||||||||||||||||||||||||||||||
Db    135906 GTTGTTGACCCAGGCGGTTAAACTTACCTCAGATTATTTTTGTCTCTGTTTGTCTTTACTGTGTC 135965

Qy      3229 AGCTACCTCAGTG-TTTTTACCTCAGCAGCATTTTTAATAAAATGAGTTGGAAAAAAAT 3288
             ||||||||||||| ||||||||||||||||||||| ||||||||||||||||||||||
Db    135966 AGCTACCTCAGTGTTGGAGCATTTTTAATAAAATGAGTTGGAAAAAAAT 136025

Qy      3289 AAGATATCAACTGCCAGCCTGGAAGTGTCGACAGTCAAGCTGCAACAGCTGTTCTGAA 3348
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    136026 AAGATATCAACTGCCAGCCTGGACGCTGGTGTTTTTCCAAATGTGCAACAGCTGTTCTGAA 136085

Qy      3349 TTGTCTTCCGCTAGCCAAGAAACCNATATGGCCTTCTTTGGACAAACCTTGAAAATGTTT 3408
             |||||||| |||||||||||||| |||||||||||||||||||||||||||||||
Db    136086 TTGTCTTCTGCTAGCCAAGAAACCTATATGACCTTTCTTTGGACAAACCTTTGAAAATGTTT 136145

Qy      3409 A 3409
             |
Db    136206 A 136206

RESULT 10
```

```
LOCUS       AL596177            91084 bp    DNA     linear   PRI 08-SEP-2001
DEFINITION  Human DNA sequence from clone RP11-325P15 on chromosome 1, complete
            sequence.
ACCESSION   AL596177
VERSION     AL596177.4  GI:15552942
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 91084)
AUTHORS     Clark,S.
TITLE       Direct Submission
JOURNAL     Submitted (08-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 11, 2001 this sequence version replaced gi:15022347.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-325P15 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-325P15 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP11-325P15 is at 1 in this sequence.
            The true left end of clone RP11-337C18 is at 89085 in this
            sequence. The true right end of clone RP4-704D21 is at 41034 in
            this sequence.
FEATURES             Location/Qualifiers
     source          1..91084
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-325P15"
                     /clone_lib="RPCI-11.2"
ORIGIN
```

```
Query Match         77.5%;  Score 2644;  DB 9;  Length 91084;
Best Local Similarity   90.0%;  Pred. No. 0;
Matches 3090;  Conservative 0;  Mismatches 166;  Indels 177;  Gaps 17;

Qy      1 GGCAGCGGAGGAGGCGAGGAGCGCGCCGGAGCCGGGGAGCCCGCGGGGCTCTCGGGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9381 GGCAGCGGAGGAGGCGAGGAGCGCGCCGGAGCCGGGGAGCCCGCGGGGCTCTCGGGG 9440

Qy     61 AAGAGACCGATGATGAACAAGCTTTACATCGGGACAAAGCTTCAGCTGGAACCTGAGC 120
          ||||||||| ||||||||   ||||||||||||  ||||||||||||||  ||||||||
Db   9441 AAGAGACCGGTGATGAACAATCTTTACATCGGGAACAATCTTTACATCGGGAACCTTAGCCTGAGC 9500
```

```
Qy    121 GACCTCCGGCAGCTCTTCTTGGGGACAGGAAGTGCCCCTGGCGGGGACCAGGTCTGCTGAAG 180
          |||||| ||||||||||  |||||||||||||||||||   ||||||||||||||||| |||
Db   9501 GACCTCTGGCAGCTCTTTGGGGACAGGAAGTGCCCCCACCGGACCAGGTCTGCTGCTCAAG 9560

Qy    181 TCCGGCTACGCCTTCGTGGACTACCCCTACAAGAACTGGGCCATCGGGCCATCGAGACC 240
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9561 TCCGGCTATGCCTTCGTGGACTACCCCTACAAGAACTGGGCCATCGGGCCATCGAGACC 9620

Qy    241 CTCTCGGGTAAAGTGGAATTGCATGGGAAATCATGGAGAGTTGATTACTCAGTCTCTAAA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9621 CTCTCGGGTAAAGTGGAATTGCATGGGAAATCATGGAGAGTTGATTACTCAGTCTCTAAA 9680

Qy    301 AAGCTAAGGAGGAGGAAAATCAGATTCGAAACATCCCTCCTCACCTGCAGTGGGAGGTG 360
          ||||||||||||||||||| |||||||||||||||||||||||||||  |||||||||||
Db   9681 AAGCTAAGGAGGAGGAAAAATCAGATTCGAAACATCCCTCCTCACCTCCAGTGGGAGGTG 9740

Qy    361 TTGGATGGACTTTTGGCTCAATATGGACAGTGGAGAATGTGG---AACAAGTCAACACA 417
          |||||||||||||||||||||||||||||||||||||         ||||||||||||||
Db   9741 TTGGATGGACTTTTGGCTCAATATGGACAGTGGAGACAACAACAAGTCAACACA 9800

Qy    418 GACACAGAAACCGCCGTTGTCAACGTCACATATGCAACAGAGGAAGAAGCAAAATAGCC 477
          ||||||||| |||||||||||||| |||||||||||||||||||     ||||||||||
Db   9801 GACACAGAAACCACCTGTGTCAACGTCACATGCAACAAGAGAAG-----AAAATAGAC 9855

Qy    478 ATGGAGAAGCTAAGCGGGCATCAGTTGAGAACTACTCCTTCAAGATTCCTACATCCCG 537
          |||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db   9856 ATGGAGAAGCTAAGCGGGCATCAGTTGAGAACTAGTCCTTCAAGATTCCTACATCCCG 9915

Qy    538 GATGAAGAGGTGAGCTCCCCTTCGCCCCCCAGCGAGCCCCAGCGTGGGGACCACTCTTCC 597
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   9916 GATGAAGAGGTGAGCTCCCCTTCGCCCCCCAGCGAGCCCCAGCGTGGGGACCACTCTTCC 9975

Qy    598 C-GGGCAAGCCACGCCTGGGGAACCCTGGGGGCCGTTCTTAGGCCAGACAGATGATTCCCGCT 656
          |  ||||||||||||| ||||||||||||| |||  ||||||||||| |||||||||||||||
Db   9976 CGGGGCAAGCCACGACCCACGCCGCGTTCTTAGGCCAGACAGATTGATTCCTGCT 10035

Qy    657 GCGGGATCCTGGTCCCCACC------CAGTTTGTTGGTGCCATCATCGGAAGGAGGGGCTT 710
          |||||| || ||||||||       |||||| ||  || ||||| |||||||||||||||
Db  10036 GCAGGTTCCTGGTCCCCTACAGTTTGCAGTTTGTTGGTGCCATCATCAGAAGAAGGAGGGGCTT 10095

Qy    711 GACCATAAAGAACATCACTAAGCAGCACACCAGTCCGGGTAGATATCCATAGAAGAAAGAGAA 770
          |||||||||||||||||||||||||| ||||||||||  ||||| |||||||| |||||||||
Db  10096 GACCATAAAGAACATCACTAGTAGACCCCAGTCCGGGATAGTATATCCATAGAAAGAGAA 10155

Qy    771 CTCTGAGCTGCAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGC 830
          ||||||||||||||||||||||||||||||||||| ||||||||| |||||||||||
Db  10156 CTCTGAGCTGCAGAAGCCTGTCACCATCCATGCCACCCCAGGGGGACTTCTGAAGC 10215

Qy    831 ATGCCCATGATTCTTGAAATCATGCAGAAGAGGCAGATGAGACCAAACTAGCCGAGA 890
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db  10216 ATGCCCATGATTCTTGAAATCATGCAGAAGAGGCAGATGAGATGAGAACAAACTAGCCGAAGA 10275

Qy    891 GATTCCTCTGAAAATCTTGGCACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGG 950
          ||||||||||||||  |||||||  ||||||||||||       ||  |||||||| ||
Db  10276 GATTCCTCTGAAAATCTTGGCCCACACTGGTTGG--------TTGGAAAAGAAGG 10323

Qy    951 CAGAAATTTGAAGAAATTGAACATGAAACATGAAGAGCTGCTGAGAATGCATCTCTTTGCA 1010
          ||||||||||||||||||||||||||||||| |||||||||||||||| ||||| ||||||
Db  10324 CAGAAATTTGAAGAAACTGAACAATAACAAGAAGATAGCAAATCTCATCTTTGCA 10383

Qy   1011 GGATTTGAGCATATACAACCCGAAGAACCATCACTGTGAAGGGCACAG-TTGAGGGCCT 1069
          |||||||||||||| || ||||||||||||||||||||||||||||||  ||||||||||
Db  10384 GGATTTGAGCATAAATAACCAGGAAGAATCATCACTGTGAAGGGCACAGTTTGAGGGCCT 10443

Qy   1070 GTGCCAGTGCTGAGATAGGATTATTGAAAATGATATGC 1129
          |||||||||||||||||||||||||||||||||| |||
Db  10444 GTGCCAGTGCTGAGATAGAGGATTATTGAAGAAGCTTTGAAGAACCTTTGAAAAATGATATGT 10503

Qy   1130 TGGCTGTTAACCAACAAGCCAATCTGATCCCAGGGTTGAACCTTGCACTTTGGCATCT 1189
          |||||||||  ||||||
Db  10504 TGGCTGTTAA 10513

Qy   1190 TTTCAACAGGACTGTCTCCGTGCTATCTCCACCAGCAGGACCAGGGACCCGGGAGTCCCCCCGCTG 1249
```

```
Db  10514 -------------------------------------------------------- 10513    Db  11513 ACATATTAGAAGGGAAAGATGTTAAGAATATATGGGCCTGTGGGTTACACAGGGTGCCTGCA 11572
Qy   1250 CCCCTACCACCCCTTCACTACCTCTCCGATACTTCTCCAGCCTGTACCCCATCACC        1309    Qy   2329 GCGTAATATATATTTTAGAAATATATATATCAAATAACTCAACTCAAC-TTTAATCA       2388

Db  10514 --------------TACGGCACTCCGGATACTTCTCCAGCTTGTACCCCATCACC        10554    Db  11573 GCAGTAATAATATTTTAGAAAATATCAAATAATATCAAATAACTCATCTA----CAATTTTTAATCA 11628
Qy   1310 AGTTTGGCCCGTTCCCGCATCATCATCACTCTTATCCAGAGCAGAGATTGTGAATCTCTTCA 1369    Qy   2389 ATTATTAATTTTTTTTCTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAGATAA     2448

Db  10555 AGGTTGGCCCGTTCCCGCATCATCATCACTCTTATCCAGAGCAGAGGTTGTGAATCTCTTCA 10614    Db  11629 ATCATTAA-TTTTTTCTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAAGAGATA      11687
Qy   1370 TCCCAACCCAGGCTGTGTGGGCCCATCATCGGAAGAAAGGGGCACACATCAAACAGCTGG   1429    Qy   2449 A-----GTCTTTGGGAGGTCTCACCGTGTAGAGAGAGCTTTGAGGCCACCCGCACAAAA     2503

Db  10615 TCCCAACCCAGGCGTGTGGGCCCATTATCAGGAGGAGAGGGAGCACACATCAAACAGCTGG  10674    Db  11688 AAGTCTTTCTTTGGGAGGTCTCACACAGTATAGACAGAGCTTTGAGGCCACCCACCACAAAA 11747
Qy   1430 CGAGATTCGCCCGGAGCCTTATCAAGATTGCCCCTGCGCAGAAGGCCAGACGTCAGCCGAAA 1489    Qy   2504 TTTCACCCAGAGGGAAATCTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTA   2563

Db  10675 CGAGATTCGCCCACAGCCCTCCATCAAGATCGCCCTTGCGGAGGCCCAGACGTCAACGAAA  10734    Db  11748 TTCACCCAGAGGGAAATCTCAAAGGAAGGACAATCACAGCAGTTCTGGATCACCTGTGTA    11807
Qy   1490 GGATGGTCATCATCACCGGGAGCCACCGGAGCCCAGTTCAAGGCCCAGGACGGATCTTTG  1549    Qy   2564 TGTCAACAGAAGGGATACCGTCCTCCTTGAAGAGGAAACTCTGTCACTCCTCATGCCTGTC   2623

Db  10735 GGATGGTCATCATCACCACCCCGCCACCGGAGCCCAGTTCAAGGCCCAGGACGGATCTTTG  10794    Db  11808 TGTCAACAGAAGGGATACCATCTCCTTGAAGAGGAAACTCTGTCACTCCTCATGCCTGTC    11867
Qy   1550 GGAAACTGAAAGAGGAGAAAACTTCTTAACCCCCAAAGAGAGAAGTGAAGCTGGAGCGCATA 1609    Qy   2624 TAGCTCATCACCCCATTTCTCTTTGCTTCACAGGTTTAAACTGGTTTTTGCATACTGC     2683

Db  10795 GGAAACTGAAAGAGGAGAAAACTTCTTAAACCCCAAAGAGAGAAGTGAAGCTGGAGCCCGTA 10854    Db  11868 TAGCTCTTACACCCCATTTCTCTTTGCTTCACAGGTTTTTAAACTGGATTTTTGCATACTGC  11927
Qy   1610 TCAGAGTGCCCTCTTCCACAGCTGGCGGTGATTGGCAAAGTGGCAAGACCCGTGAACG     1669    Qy   2684 TATATAATTCTCTGTCTCCTCCTGTTTATCTCCCCCTTCTCC                    2743

Db  10855 TCAGAGTGCCCTCTTCCACAGCTGGCGGTGATTGGCAAAGTGTCAATACCTTGAATG      10914    Db  11928 TATATAATTATCCATCTCTCCTCGTTTATCTCTCTCTCCCCTCCCGTTCTTTC         11987
Qy   1670 AACTGCAGAACTTAACCAGTGCAGAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAA    1729    Qy   2744 ATCTCCATTCTTTTGAATTCCTCATCCCCATCCCGTATCTACGACCC-CCC           2802

Db  10915 AACTGCAGAACTTAACCAGTGCAGAGTCATCGTGCCTCGTGACCAAAGGCCAGATGAAA    10974    Db  11988 ATCTTCCATTCTTTTGAATTCCCTCATCCCCATCCCGTATCTCAATCTATGCACCTCAC    12047
Qy   1730 ATGAGGAAGTGATCGTCAGAATTATCGGGACACTTCTTTGCTAGCCAGACTGCACAGCGCA  1789    Qy   2803 CCCCCCCCAGGCAAAGGCAAGAGTGCTCTGAGTATCACATCACACAAAGGAACAAAAGCGAAA 2862

Db  10975 ATGAGGAAGTGATCGTCAGAATTATTGGACACTTCTTTGCTACCCCAACTGCATAGCACA  11034    Db  12048 TCCCTCACAGGGAAAGCAGCAGTGCTCTGAGCATCGGCATCCCACCACCACGAAGGAACAAAAGTGAAA 12107
Qy   1790 AGATCAGGGAAATTGTACACAGGTGAAGCAGCAGGAGCAGAAAATACCCTCAGGGAGTCG  1849    Qy   2863 CACACAAACCAGCTCAACTTACACTTGCTTACTCTTGGTTACTCTCAAAGAGAGTCAATGGTACTT 2922

Db  11035 AGATCAGGGAAATTGTACACAGGTGAAGCAGCAGCAGAAGAGCAGAAAATACCCTCAGGGAGTCG 11094    Db  12108 CACACAAACCAACCAACCTCAATTTACACTTGGTTACTCTCAAAGAGAGTCAATGGTACAT  12167
Qy   1850 CCTCACAGCGCCAGCAAGTGAGGCTCCCAGAGGCACCACCAGCAACCAGCAACAAACAAACAAACAACA 1909    Qy   2923 GTCCTAGCGTTTGAGAGAGAAACAGAGAAAACAGAACCCACCAACAATCAACCAAACAA      2982

Db  11095 CCTCACAGCGCCAGCAAGTGAGGCTCCCAGAGGCACCACCAGCAACCAGCAACAAACAACA 11154    Db  12168 GTCCTAGCGTTTTGGAAGAGAAACAGAGAAAACAGAACCCATCAACAATCAACCAACCAACCA  12227
Qy   1910 CCTTCCAACACCTGACAGCGAGAATGAGAGAGAGAATGAGCGGAGTCGGAGCGAGCAAAACCAAA 1969    Qy   2983 AGAAAAAATTCCACAATGAAAGAATGATTGTC-TTTTGTCATTTGGTGTTATAAGGC        3041

Db  11155 CCTTCCAACAACCTGACACCTGACAGCGAGAATGAGAGAGTCGGAGCGAGCAAAACCAAA   11214    Db  12228 AGAAAAAATTCCACAATGAAAGAATGATTGTTTGTCATTTGTGTGTATAAGGC          12287
Qy   1970 GACCATCTGAGGAATGAGAGTCTGGAGGCCGGGCGGGCAGGCCCTGAGA             2029    Qy   3042 ATCAATATTCAGCAAAATGATTCTTCTTAAAAAAAATGTGGAGGAAGTAGAAA          3101

Db  11215 GACCATCTGAGGAATGAGAGTCTGCGGAGCTGGCGCACGGGGGCTGGCGCCCCTGAGA    11274    Db  12288 ATCAATATTCAGCAAAATGATTTCTTTCTTTT---AAAAAAATGTGGAGGAAGTAGAAA    12344
Qy   2030 ACCCCAGGGGCCGAGGAGGGGCGGGGAAGGTGGGGAAGGTTTTGCCAGAGCCC         2089    Qy   3102 TTTACCAAGGTTGTTGGCCCAGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAAAAAC   3161

Db  11275 ATTCCAGGGGCCGCAGAGAGAGAGGGGGTGGGGAAGGTTGGGGAAGGTTTTGCCAGAACCACTGGGCCC 11334    Db  12345 TTTACCAAGGTTGTGGCCCAGGGGGTCATTAAATTTACAGATTTTTTAATGGGAAAAAAAAC 12404
Qy   2090 -CGCCTCCCCGCCCCCCCCCCATCAGGGCTTCTGCAGGCTTGCAGCCTTCAGC          2148    Qy   3162 ACAGAAGAGCTACCTCAGTG-TTTTACCTCAGCACCTTGCTCTTGTTCTTCCCTTAG       3220

Db  11335 GCTTCTCACCCCATCAGGGCTTCTGCAGGCTTGCAGCCTTCACCATCCACTCCG         11394    Db  12405 ACAGAAAAAGCTGCCTCAGTGTTTTTACCTGCCTCAGCACCTTGCTCTTGTTCCCTTAG    12464
Qy   2149 ATCTCTCCTGAACTCCCAGACGGCCATCCCTTTGAACTAACATAGGTGAACGTGT        2208    Qy   3221 AGATTTTGTAAAGCTGATAGTTGGAGCATTTTTTA-TTTTTTAATAAAATGAGTTGG      3279

Db  11395 ATCTCCTCCTTAACTCCCCAGACGCGCATCCCTTTAGTTGAACTAACATAG--TGATGCGT  11452    Db  12465 AGATTTTGTAAAACTGATAGTTGGAGCATTTTTTAATTTTTTAATAAAATTAGTTGG      12524
Qy   2209 TCAAAGCCAAGCAAGCAAAATGCACACCCTTTTTTCTCTGTGCAAATCGTCTCTACATGTGT 2268    Qy   3280 AAAAAAAAAATAAGATAAGATCAACTGCCAGCCTGGAGAGGGTGCAAGTGTCAACAGC      3339

Db  11453 TCAAAGCCAAGCAAGCAAAATGCACACCCCTTTTTTTGTGACAAGTCATCTCTGTACATGTGTAT 11512    Db  12525 AAAAAAAATAAGATAAGATCAACTGCCAGCCTGGAGAGGTGCAAGTGTGCAACAGC       12584
Qy   2269 ACATATTAGAAAGGGAAGATGTTAAGAATATATGGGCCTGTGGGTTACACAGGGTGCCTGCA 2328    Qy   3340 TGTTCTGAATTGTCTTCTCCGTCCGTCTGCTAGCCAAGAACCNTATATGGCCTTCTTTGGCAAAACCTTG 3399

                                                                                   Db  12585 TGTTCTGAATTGTCTTCTGCTAGCCAAGAACCTTATATGGCCTTCTTTGGCAAAACCTTG   12644
```

```
Qy          3400  AAAATGTTTATTT 3412
                  ||||||||||||
Db         12645  AAAATGTTTATTT 12657
```

```
RESULT 11
AC015706
LOCUS         AC015706          182695 bp    DNA     linear    HTG 24-AUG-2002
DEFINITION    Homo sapiens chromosome 1 clone RP11-325P15 map 1, WORKING DRAFT
              SEQUENCE, 17 unordered pieces.
ACCESSION     AC015706
VERSION       AC015706.3  GI:8096833
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 182695)
AUTHORS       Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE         Homo sapiens chromosome 1, clone RP11-325P15
JOURNAL       Unpublished
REFERENCE     2  (bases 1 to 182695)
AUTHORS       Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
              Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
              Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
              Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
              Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
              Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
              Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
              Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
              McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
              Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
              Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
              Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
              Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
              Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE         Direct Submission
JOURNAL       Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE     3  (bases 1 to 182695)
AUTHORS       Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
              Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
              Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
              Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
              Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
              Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
              Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
              Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
              Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
              Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
              Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
              McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
              Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
              Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
              O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
              Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
              Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
              Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
              Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
              Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
              Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE         Direct Submission
JOURNAL       Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT       On May 26, 2000 this sequence version replaced gi:6454072.
              All repeats were identified using RepeatMasker:
              http://ftp.genome.washington.edu/RM/RepeatMasker.html
              Smit, A.F.A. & Green, P. (1996-1997)
              -------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
```

```
                Contact: sequence_submissions@genome.wi.mit.edu
              -------- Project Information
                Center project name: L1382
                Center clone name: 325_P15
              -------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 141788 bases at least Q40
                Consensus quality: 163541 bases at least Q30
                Consensus quality: 174666 bases at least Q20
                Insert size: 181000; agarose-fp
                Insert size: 181095; sum-of-contigs
                Quality coverage: 4.2 in Q20 bases; agarose-fp
                Quality coverage: 4.2 in Q20 bases; sum-of-contigs

              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 17 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *        1      2035:  contig of 2035 bp in length
              *     2036      2135:  gap of 100 bp
              *     2136      6069:  contig of 3934 bp in length
              *     6070      6169:  gap of 100 bp
              *     6170     10460:  contig of 4291 bp in length
              *    10461     10560:  gap of 100 bp
              *    10561     14300:  contig of 3740 bp in length
              *    14301     14400:  gap of 100 bp
              *    14401     18480:  contig of 4080 bp in length
              *    18481     18580:  gap of 100 bp
              *    18581     23093:  contig of 4513 bp in length
              *    23094     23193:  gap of 100 bp
              *    23194     28213:  contig of 5020 bp in length
              *    28214     28313:  gap of 100 bp
              *    28314     32808:  contig of 4495 bp in length
              *    32809     32908:  gap of 100 bp
              *    32909     37000:  contig of 4092 bp in length
              *    37001     37100:  gap of 100 bp
              *    37101     45183:  contig of 8083 bp in length
              *    45184     45283:  gap of 100 bp
              *    45284     53421:  contig of 8138 bp in length
              *    53422     53521:  gap of 100 bp
              *    53522     64476:  contig of 10955 bp in length
              *    64477     64576:  gap of 100 bp
              *    64577     75144:  contig of 10568 bp in length
              *    75145     75244:  gap of 100 bp
              *    75245     93230:  contig of 17986 bp in length
              *    93231     93330:  gap of 100 bp
              *    93331    116951:  contig of 23621 bp in length
              *   116952    117051:  gap of 100 bp
              *   117052    143611:  contig of 26560 bp in length
              *   143612    143711:  gap of 100 bp
              *   143712    182695:  contig of 38984 bp in length
FEATURES             Location/Qualifiers
     source          1..182695
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1"
                     /clone="RP11-325P15"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..2035
                     /note="assembly_fragment"
     misc_feature    2136..6069
                     /note="assembly_fragment"
     misc_feature    6170..10460
                     /note="assembly_fragment"
     misc_feature    10561..14300
```

```
misc_feature   14401..18480
               /note="assembly_fragment"
misc_feature   18581..23093
               /note="assembly_fragment"
misc_feature   23194..28213
               /note="assembly_fragment"
misc_feature   28314..32808
               /note="assembly_fragment
               clone end:T7
               vector side:right"
misc_feature   32909..37000
               /note="assembly_fragment"
misc_feature   37101..45183
               /note="assembly_fragment"
misc_feature   45284..53421
               /note="assembly_fragment
               clone end:SP6
               vector side:left"
misc_feature   53522..64476
               /note="assembly_fragment"
misc_feature   64577..75144
               /note="assembly_fragment"
misc_feature   75245..93230
               /note="assembly_fragment"
misc_feature   93331..116951
               /note="assembly_fragment"
misc_feature   117052..143611
               /note="assembly_fragment"
misc_feature   143712..182695
               /note="assembly_fragment"

ORIGIN

Query Match       77.4%;   Score 2642.4;   DB 2;   Length 182695;
Best Local Similarity  90.0%;   Pred. No. 0;
Matches 3089;   Conservative  0;   Mismatches  167;   Indels 177;   Gaps  17;
```

```
Qy       1 GGCAGCGGAGGAGGAGGCGGCCGGTACCGGCGCCGGGGAGGCGCGGGGAGGCGCCGGGGAGGCGCCGGGCGGTCTCGGGGGG 60
           ||||||||||||||||||||||| |||||||||||||||| ||||||||||| |||||||
Db   19246 GGCAGCGGAGGAGGAGGCGGCCGGTACCGGCGCCGGGGAGGCGCGGGGAGGCGCCGGGGAGGCGCCGGGGAGGCGCCGGGGG 19305

Qy      61 AAGAGACGGATGATGAACAAGCTTTACATCGGGAACAAGC 120
           ||||||||||||||||||||||||||||||||||||||||
Db   19306 AAGAGACGGGTGATGATGAACAATCTTTACATCGGGAACAAGC 19365

Qy     121 GACCTTCCGGCAGCTCTTTGGGGACAGGAAGCTGCCCTGCGCGGACAGGTCCTGCTGAAG 180
           ||||||||||||||||||||||||||||
Db   19366 GACCTTCCGGCAGCTCTTTGGGGACACGGAAGCTGCCCCACCGGACAGGTCCTGCTCAAG 19425

Qy     181 TCCGGGTACCGGCTGGGGACTACCCCGACCGGACCGGACC 240
           ||||||||||||||||||||
Db   19426 TCCGGGTATGCCTTGCTGGGGACTACCCCTACACAGTACTGGGCCGGACTCGGGCCATCGGGAAATCATGGAAGTTGATTACTCAGACC 19485

Qy     241 CTTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTTCAGTTCTAAA 300
           ||||||||||||||||||||||||||||||||
Db   19486 CTTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTTCAGTTCTAAA 19545

Qy     301 AAGCTAAGGAGCAGGAGAAAATTCAGATTGGGACAGTGGG 360
           ||||||||||||||||||||||||||||
Db   19546 AAGCTAAGGAGCAGGAGAAAAATTCAGATTGGGACAGTGGGAGGTG 19605

Qy     361 TTGGATGGACTTTTGGCTCAATATGGGACACAGTGGG 417
           |||||||||||||||||||||||||||||||||
Db   19606 TTGGATGGACTTTTGGCTCAATATGGGACACAGTGGGAACAAGTCAACACA 19665

Qy     418 GACACAGAGAACCACTGTGTCAAGGTCACAGTCACAGTCACACAGAGAGAAG 477
           |||||||||||||||||||||||||||||||||
Db   19666 GACACAGAGAACCACTGTGTCAAGCGTCACAGTCACAGTCACAAGAGAGAAG 19720

Qy     478 ATGGAGAGCTAAGCGGCATCAGTTGCAGAACTTGAGAAGTTTGAGAACTTGAGAACTACTCCTTCAAGAT 537
           ||||||||||||||||||||||||||||||||||||||||
Db   19721 ATGGAGAGCTAAGCGGCATCAGTTGCAGAACTAGTCCTTCAAGATTTCCTACATCCCG 19780
```

```
Db  20720  |||||||||||||||||||||||||||||||||||||  |||||||||  ||  ||||
    20779
Qy   1670  AACTGCAGAACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAA  1729
Db  20780  AACTGCAGAACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAA  20839
Qy   1730  ATGAGGAAGTGATCGTCAGAATTATCGGGACACTTCTTCGTAGCTGAGCGACGCACGCCA  1789
Db  20840  ATGAGGAAGTGATCGTCAGAATTATTGGACACTTCTTCGTACCCCAACTGCATAGCACA  20899
Qy   1790  AGATCAGGGAAATTGTACAACAGTGAAGCAGTGACCAGAAATACCCTCAGGGAGTCG  1849
Db  20900  AGATCAGGGAAACTGTACAACAGTGAAGCAGCAGAGACAGAAATACCCTCAGGGAGTCG  20959
Qy   1850  CCTCACAGACGCAGCAAGCAAGTGAGGCTTCCCACAGGCACCAGAGCCACCAGGAATGTAGC  1909
Db  20960  CCTCACAGCACAGCAGCAAGTGAGGCTTCCCACAGGACCAGAGCCACCAGGAATGTAGC  21019
Qy   1910  CCTTCCAACACCTGACAGAATGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAA  1969
Db  21020  CCTTCCAACACCTGACAGAATGAGACCAAACAGCCAGCCAGATCGGGAGCAAACCAAA  21079
Qy   1970  GACCATCTGAGGGAATGAGAAGTCTGCGAGGCGGCCAGGGACTCTGCCGAGGCCCTGAGA  2029
Db  21080  GACCATCTGAGGGAATGAGAAGTCTGCGGAGGCGGCTCCCACGGACTCTGAAGAGGCCCTGAGA  21139
Qy   2030  ACCCGAGGGGCGGAGGAGGGCGGGGAAGTCAGCCAGGTTCTGCAGAACCACCGGAGCCC  2089
Db  21140  ATTCCAGGGGGCGGAGGAGGGGGTGGGGAAGTCAGCCAGGTTGCCAGAACCACTGGGGCCC  21199
Qy   2090  -CGCCTCCGCGCCCCCCCCCATCCAGGGCTTCTGCAGGCTTCACCTTCACCATCCACTCGG  2148
Db  21200  GCTTCTCACCCCCATCCAGGGCTTCTGCAGGCTTCAGCAGGCTTCACCATCCACTCCG  21259
Qy   2149  ATCTCTCCTGAACTCCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGT  2208
Db  21260  ATCTCTCCTTAACTCCCACGACGCTATCCCTTTTAGTTGAACTAACATAG--TGATGCGT  21317
Qy   2209  TCAAAGCCAAGCAAAATGCACACCCTTTTTCTGTGGCAAATCGTCTCGTACATGTGTGT  2268
Db  21318  TCAAAGCCAAGCAAAATGCACACACCCTTTTTTGTGACAAGTCATCTCGTACATGTGTAT  21377
Qy   2269  ACATATTAGAAGGGAAGATGTTAAGATATGTGCCTTGTGTGGGTTACACAGGGTGCTGCA  2328
Db  21378  ACATATTAGAAGGGAAGATGTTAAGATGTGCCTTGTGTGGGTTACACAGGGTGCTGCA  21437
Qy   2329  GCGGTAATATATATTTAGAAAATAATAATATCAAATAATCAACTAACTCCAATTTTAATCA  2388
Db  21438  GCAGTAATATATATTTAGAAATAATAATATCAAATAATCAAACTAACTCATCTA----CCATTTTAATCA  21493
Qy   2389  ATTATTAATTATTTTTTCTTTTTCTTTTCTTTTAAAGAGAAAGCAGGCTTTTCTAGAATAA  2448
Db  21494  ATCATTAA-TTTTTTTCTTTTTCTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAAGAGAATA  21552
Qy   2449  A-----GTCTTTGGGAGGTCTCACGGGTGTAGAGAGAGGCTTTGAGGCCACCCGCACAAAA  2503
Db  21553  AAGTCTTTTCTTTGGGAGGTCTCACAGTATAGACAGAGGCTTTGAGGCCACCCACACAAAA  21612
Qy   2504  TTCACCCGAGGGGAAATCTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTA  2563
Db  21613  TTCACCCCAGAGGGAAATCTCAAAGGAAGGACAATCACGCAGTTCTGGATCACCTGTGTA  21672
Qy   2564  TGTCAACAGAAGGGATACCGTCCGTCCTCCTTGAACAGGAGAAACTCTGTCACTCCTCATGCCTGTC  2623
Db  21673  TGTCAACAGAAGGGATACCATCCCTCCTTGAAGAGGAGAAACTCTGTCACTCCTCATGCCTGTC  21732
Qy   2624  TAGCTCATACACCCATTTCTTCTTGCTTGCTTCCAGAGTTTAAACTCTGTTTTTTGCATACTGC  2683
Db  21733  TAGCTCTTACACCCATTTCTTCTTGCTTTGCTTCCAGAGTTTTAAACTGGATTTTTTGCATACTGC  21792
Qy   2684  TATATAATTCTGTCTCTGTTTATCTCGTTAATCTCTCCCCTTCTCC  |||||||  ||||||  2743
```

```
Db  21793  TATATAATTATCCATCTCTCTGTTTATCTCCCCCTCCCCCGTCTTCTTC  21852
Qy   2744  ATCTCCATTCTTTTGAATTCCCATCCCTCCATCTCAATCCCGTATCTAGCGCACC-CCC  2802
Db  21853  ATCTCCATTCTTTTGAATTCCCATCCCTCCATCTCAATCCCGTATCTATCGCACCTCAC  21912
Qy   2803  CCCCCCCAGGCAAAGCAGTGCTCTGAGTATCACATCACACAAAAGGAACAAAAGCGAAA  2862
Db  21913  TCCCTCACAGGGAAAGCAGTGCTCTGAGTATCGCCATCACCAGCAAAGGAACAAAAGTGAAA  21972
Qy   2863  CACACAAACCAGCCTCAACTTACACTTGGTTACTCACACAGAGAACAAGAGTCAATGGTACTT  2922
Db  21973  CACACAAACCAACCTCCAATTTACACTTGGTTACTCACAAAGAAGAACAAGAGTCAATGGTACAT  22032
Qy   2923  GTCCTAGCGTTTTGGAAGAGGAAAACAGGAAACCACCAGAGCCACCAAACCAATCAACCAAAACAA  2982
Db  22033  GTCCTAGCGTTTTGGAAGAGGAAAACAGGAAACCAACCAATCAACCAATCAACCAAACAA  22092
Qy   2983  AGAAAAAATTCCACAATGAAAGAATGTATTTTGTC-TTTTTGCATTTTGTGTATAAGCC  3041
Db  22093  AGAAAAAATTCCACAATGAAAGAAGAGGTGTTTTGTCTTTTTGGTATTTTGTGTATAAGCC  22152
Qy   3042  ATCAATATTCAGCAAAATGATTCCTTTCTTTAAAAAAAAATGTGGAGAAAGTAGAAA  3101
Db  22153  ATCAATATTCAGCAAAATGATTTCTTTCTTT---AAAAAAAATGTGGAGAAAGTAGAAA  22209
Qy   3102  TTTACCAAGGTTGTTGGCCCAGGGCGTTAAATTCACAGATTTTTTAACAGAGAAAAACAC  3161
Db  22210  TTTACCAAGGTTGCTGCCCAGGGCCATTAAATTTTACAGATTTTTTAATGGGAAAAAACAC  22269
Qy   3162  ACAGAAGAAGTCTACCTCAGGTG-TTTTTACCTCAGCACCTTGTCTCTTGTCGTTCCCTTAG  3220
Db  22270  ACAGAAGAAAGCTGCCTCAGGGTGTTTTTTTACCTCAGCACCTTGTCTCTTGTCGTTCCCTTAG  22329
Qy   3221  AGATTTTGTAAAGCTGATAGTTGGAGCATTTTTTA-TTTTTTTATAAAATGAGTTGG  3279
Db  22330  AGATTTTGTAAAAACTGATAGTTGGAGCATTTTTTAATTTTTTTAATAAAAATTAGTTGG  22389
Qy   3280  AAAAAAATAGATATCAACTGCCAGCCTGGAGAGGTGACAGTCCAAGTGTGCAACAGC  3339
Db  22390  AAAAAAATAGATATCAACTGCCAGCCTGGAGAGGTGACAGTCCAAGTGTGCAACAGC  22449
Qy   3340  TGTTCTGAATTGTCTTCCGCTAGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTG  3399
Db  22450  TGTTCTGAATTGTCTTCTGCTAGCCAAGAACCTATATGGCCTTCTTTTGGACAAACCTTG  22509
Qy   3400  AAAAATGTTTATTT  3412
Db  22510  AAAAATGTTTATTT  22522
```

```
RESULT 12
AC104980
LOCUS      AC104980  187226 bp  DNA  linear  PRI 31-MAR-2002
DEFINITION Homo sapiens chromosome 8, clone RP11-281D17, complete sequence.
ACCESSION  AC104980
VERSION    AC104980.5  GI:19849375
KEYWORDS   HTG.
SOURCE     Homo sapiens (human)
  ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 187226)
  AUTHORS  Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE    Homo sapiens chromosome 8, clone RP11-281D17
  JOURNAL  Unpublished
REFERENCE  2 (bases 1 to 187226)
  AUTHORS  Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
           Anderson,S., Barna,N., Bastien,V., Bogdanovskiy,L., Boukhgalter,B.,
           Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
           Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
           Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
           Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,D.,
```

```
          Smit,A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          ---------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence submissions@genome.wi.mit.edu
          --------- Project Information
            Center project name: L22878
            Center clone name: 281_D_17
          ---------
FEATURES             Location/Qualifiers
     source          1..187226
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-281D17"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   1..5915
                     /rpt_family="L1PA5"
     repeat_region   5916..5943
                     /rpt_family="AT_rich"
     repeat_region   5944..6477
                     /rpt_family="t1"
     repeat_region   8563..8886
                     /rpt_family="MER2"
     repeat_region   complement(9296..9555)
                     /rpt_family="L2"
     repeat_region   complement(9865..10094)
                     /rpt_family="MIR"
     repeat_region   complement(10346..10504)
                     /rpt_family="AluSg/x"
     repeat_region   complement(10821..11111)
                     /rpt_family="AluSc"
     repeat_region   11162..11300
                     /rpt_family="MER47A"
     repeat_region   11301..11596
                     /rpt_family="AluSx"
     repeat_region   11597..11836
                     /rpt_family="MER47A"
     repeat_region   12185..12319
                     /rpt_family="FLAM_C"
     repeat_region   12520..12822
                     /rpt_family="AluSp"
     repeat_region   complement(12885..13167)
                     /rpt_family="AluJb"
     repeat_region   complement(13230..13379)
                     /rpt_family="MIR"
     repeat_region   complement(14179..14479)
                     /rpt_family="AluSg"
     repeat_region   complement(14630..14747)
                     /rpt_family="MIR"
     repeat_region   14748..15055
                     /rpt_family="AluSq"
     repeat_region   complement(15056..15145)
                     /rpt_family="MIR"
     repeat_region   complement(15206..15400)
                     /rpt_family="MIR"
     repeat_region   15576..15606
                     /rpt_family="AT_rich"
     repeat_region   complement(15607..15901)
                     /rpt_family="AluSp"
     repeat_region   16127..16556
                     /rpt_family="L1ME4A"
     repeat_region   16657..16684
                     /rpt_family="AT_rich"
     repeat_region   complement(17149..17513)
                     /rpt_family="MLT1A1"
     repeat_region   17966..18067
                     /rpt_family="MER47B"
     repeat_region   complement(18068..18365)
```

```
          Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
          Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
          Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
          Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
          MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
          McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
          Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
          Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
          Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
          Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
          Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
          Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
          Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
          Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
          Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
          Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL Submitted (22-DEC-2001) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3  (bases 1 to 187226)
  AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
          Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
          Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
          Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
          Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
          Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
          Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
          Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
          Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
          Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
          MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
          McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
          Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
          Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
          Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
          Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
          Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
          Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
          Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
          Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
          Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
          Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 4  (bases 1 to 187226)
  AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
          Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
          Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
          Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
          Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
          Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
          Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
          Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
          Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
          Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
          MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
          McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
          Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
          Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
          Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
          Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
          Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
          Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
          Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
          Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
          Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
          Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL Submitted (31-MAR-2002) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Mar 31, 2002 this sequence version replaced gi:19683374.
          All repeats were identified using RepeatMasker:
```

```
repeat_region    /rpt_family="AluJo"
                 18366..18666
repeat_region    /rpt_family="MER47B"
                 19538..19561
repeat_region    /rpt_family="AT_rich"
                 complement(19562..21070)
repeat_region    /rpt_family="L1PA2"
                 complement(21072..21141)
repeat_region    /rpt_family="U3"
                 22731..22771
repeat_region    /rpt_family="MIR"
                 complement(23445..23902)
repeat_region    /rpt_family="L1PA5"
                 complement(25000..25014)
repeat_region    /rpt_family="FLAM_C"
                 25015..25083
repeat_region    /rpt_family="(TTCA)n"
                 complement(25084..25200)
repeat_region    /rpt_family="FLAM_C"
                 25393..25450
repeat_region    /rpt_family="AT_rich"
                 complement(25762..26547)
repeat_region    /rpt_family="L1ME4A"
                 26829..26889
repeat_region    /rpt_family="(TAGA)n"
                 27373..27795
repeat_region    /rpt_family="LTR37A"
                 complement(29177..29247)
repeat_region    /rpt_family="MIR"
                 29248..29370
repeat_region    /rpt_family="AluJo"
                 29371..29395
repeat_region    /rpt_family="(CA)n"
                 29396..29562
```

```
Query Match         64.2%;  Score 2189.6;  DB 9;  Length 187226;
Best Local Similarity  85.8%;  Pred. No. 0;
Matches 2727;  Conservative  0;  Mismatches  210;  Indels 242;  Gaps  17;
```

```
Db  80110 CACTGAGCAGACCCAGTCCCGGTAGACATCCATCAGAAAAGAGAACTCTGGAGCTGCAGA 80169

Qy    786 GAAGCCTGTCACCATCCATGCCACCCCAGAGGGACTTCTGAAGCATGCCGATGATTCT 845
          ||||||||||||||||||||| |||||||||||||||||||| |||||||||
Db  80170 GAAGCTTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCACCACGATTCT 80229

Qy    846 TGAAATCATGCAGAAAATTCAGAATCAGAGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAAT 905
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db  80230 TGAAAGCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAAT 80289

Qy    906 CTTGGCACACAATGGCTGGTTGGAAGACTGATTGGAAAGAAGGCAGAAATTTGAAGAA 965
          ||||||||||||||||||||| |||||| |||||||||||||| |||||||||
Db  80290 CGTGGCCCACAATGGCTTGTTGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATGTGAAGAA 80349

Qy    966 AATTGAACATGAAAACAGGGACCAAGATAACAATCTCATCTTGCAGGATTTGAGCATATA 1025
          ||||||||||||||| ||||||||||||||||||||||||||||||||||
Db  80350 AAATGAACATGAAAACAGGGACCAAGATAACAATCTCATCCTGCAGGATTTGAGCATATG 80409

Qy   1026 CAACCCGGAAAGAACCATCACTGTGCGTGAGGGCACAGTTGAGCCTGTGCCAGTGCTGAGAT 1085
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  80410 CAACCCGGAAAGAACCATCACTGTGCAGGGCCACAGTTGAGGGCCATGCCAGTGCTGAGAT 80469

Qy   1086 AGAGATTATGAAGAAGCTGCGTGGAGCCTTTGAAAATGATATGTCTGGCTGTGTTAACCAACA 1145
          ||||||||||||||||||||||||||||||||||||||||||
Db  80470 GGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGTGGCTGTGTTAA----- 80523

Qy   1146 AGCCAATCTGATCCCAGGGTTGAACCTGCACTTGGCATCTTTTCAACAGGACTGTC 1205
                                                           |||||||
Db  80524 -------------------------------------------------------- 80523

Qy   1206 CGTGCTATCTCCACCAGCAGCAGGGCCCCGCGGAGCTCCCCGCTGCCCCCTACCACCCCTT 1265
                                                           |||||||
Db  80524 -------------------------------------------------------- 80523

Qy   1266 CACTACCCACTCCGGATAC-TTCTCCAGCGTGACCCTGTACCCCCATCACCAGTTTGCCCGTTCC 1324
          ||||||||||||||||||  |||||||||||||||||||||||||||||||||||||||
Db  80524 ---TACCCACTCCAGATACTTCTCCAGCGTGACCCTGTACCCCCATCACCAGTTTGCCCGTTCC 80580

Qy   1325 CGCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTTCTTCATCCCAACCCAGGCTG 1384
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  80581 CGCATCAGCATTCTTATCGGGAAAATCATGGAAGTTGATTCTTCATCCCAACCCAGGCTG 80640

Qy   1385 TGGGGCCCATCATCGGGAAGAACATCCCTCCTCACCTGCAGTGGAGGTGTGTGGAT 1444
          ||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db  80641 TGGGTACCCATCATCGGGAAGAACATCCCTCCTCACCTGCAGTGGGCCGATGTGTTGGAT 80700

Qy   1445 CCTCTATCAAGATTGCCCCTGCGGAGATGTGGAACAAGTCAACACAGCAACAGAA 1504
          ||||||||
Db  80701 CCTCCA------------------------------------------------ 80706

Qy   1505 CCGGGCACCGGAAGCCCAGTTCAAGGCCCAGGACGGATCTTTGGGAAACTGAAAGGAGAAG 1564
                                                ||||||||||||||||||||
Db  80707 ----------------------------------TCAAGGCCCAGGACGGATCTTCGGGAAACTAAAAGAG 80745

Qy   1565 AAAACTTCTTTAACCCCAAGAAGAAGTTGAAGGCCATATCAGAGTGCCCTCTT 1624
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  80746 AAAACTTCTTTAACCCCAAGAAGAAGTTGAAGGGCCCCATATCAGAGTGCTCTCTT 80805

Qy   1625 CCACAGCTGGCCCGGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTGCAAACTTAA 1684
          |||||||||||||||||||||||||||||||||||||||||||||||
Db  80806 CCACAGCTGGCCCGGGTGATTGGCAAAGGTGGCAAAGTGCATAACTTAA 80865

Qy   1685 CCAGTGCAGAAGTCATCGTGCCTCGTGACCAGATTGATTTCCCGCTGCGGATCCT 1744
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  80866 CTAGTGCAGAAATCATTGTGCCTCGTGACCAGACTGATTTCCCGCTGCGGATCCA 80925

Qy   1745 TCAGAATTATCGGGCACTTCTTTGCTGGCCATCATCGGAATGGCCTTGACCATAAAGAAACAT 1804
          |||||||  ||||||||||||||||||||||||||||||||||||||||||
Db  80926 TCAGAATTATCGGGCACTTCTTTGCTGGCCAGTGCCAGAGTGGGCCTTGACCATAAAGAAACAT 80985

Qy   1805 TACAACAGGTGAAGCAGGAGCAGAGATATCCATAGAAAAGAGAAACTCTGGAGCTGCAGCA 1864
          ||||||||||||||||||||||||||||||||||||||||||
Db  80986 TACAAGAGGTGAAGCAGGAGCAGAGTAGAAAATACCCTCAGGGAGCTGCGCTCACAGCGCAGCA 81045
```

```
Qy    247 GGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTTAAAAGCTA 306
          |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db  79640 GGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCGGTCTCTAAAAAGCTA 79699

Qy    307 AGGAGCAGGAAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGGAGGTGTGTGGAT 366
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db  79700 AGGAACAGGAAAATTCAGATTTGAAACATTACCTCCTCCACTGCAGTGGGCCGATGTGTTGGAT 79759

Qy    367 GGACTTTTGGCTCAATATGGACAGTGGAGAATGTGGAACAAGAACAGCAACAGAA 426
          |||||||||||||| |||||||||||||||||||||||||||||||||||
Db  79760 GCACCTTTTGGCTCAACAGCGGGATAGTGGAGAATGTGGAACAAGAACAGCAACAGAA 79819

Qy    427 ACCGCCGTTGTCAACGTCACATATGCAACAAGAGAACAAGAGAACAAAAATAGCCATGGAGAAG 486
          ||||||||||||||||||||||||||||||||||||| ||||||||||
Db  79820 ACTGCTGTTGTCAATGTCACATATGCAACAGGGAAGAAGC-AAAATAGCCCGACTAAAAGAGAAG 79878

Qy    487 CTAAGCGGGGCATCAGTTTGAGAACTACTCCTCCTTCAAGATTCCCTACATCCCGATGAAGAG 546
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db  79879 CTAAGCGGGGCATCAGTTTGAGAACTACTCCTCCTTCAAGATTCCTACATCCTGGATGAAGAG 79938

Qy    547 GTGAGCT-CCCCTTCGCCCCCTCGGAGCCCAGCGTGGGGACCCACTCTTCCCGGGAGCA 605
          ||||    ||||||||||||||||||||||||||| |||||| |||||||||||||||
Db  79939 GTGAGCTCCCCCCTTACCCCCTCGGAGCCCCAGGGTGGGGACCACCCACTCTT-CCGGGGAGCA 79997

Qy    606 AGGCCACGCCCCTGGGGGCACTTCTCAGGCCAGAGCAGATTGATTTCCCGCTGCGGATCCT 665
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  79998 AGCTGCACC-CCCTGGGGGCTCTTCTCAGGCCCGACACGCTGATTTCCCGCTGCGGATCCA 80056

Qy    666 GGTCCCACCAGTTGTTGTGGCCATCATCGGAATGGCCTTGACCATAAAGAAACAT 725
          ||||                  |||||||||||||||||||||||||||||||
Db  80057 ------ACCCAGTTTGTTGGTGCCATGGTCTGAAGAAGGAGGGCTTGAACATCAGGGAAACTG 80109

Qy    726 CACTAAGCAGACCCAGTCCCGGGTAGTATATCCATCAGAAAAGAGAACTCTGGAGCTGCAGA 785
          |||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  80986 TACAAGAGGTGAAGCAGGAGCAGAGTAGAAAATACCCTCAGGGAGCTGCGCTCACAGCGCAGCA 81045
```

```
Qy   2934  TTGGAAGAGGAAAACAGGAACCCACCACCCAAAAACCAATCAACCACCAATCAACCAAACAAAGAAAAATTC  2993
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82116  TTGGAAGAGGAAAACAGGAACCCACCACCCAAAAACCAATCAACCACCAATCAACCAAACAAAGAAAAATTC  82175

Qy   2994  CACAATGAAAGAATGTATTTTGTCATTTTGTCATTTTGTGTATAAGCCATCAATATTCAG  3053
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82176  CACAATGAAAGAATGTATTTTGTCATTTTGTCATTTTGTGTATAAGCCATCAATATTCAG  82235

Qy   3054  CAAAATGAATTCCTTTTCTTAAAAAAAAAAATGGGAGGAAAGTAGAAATTTACCAAGGTT  3113
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82236  CAAAATATTTCTTTTCTTTTTTTAAAAAAATGTGAAAAATGTAGCAATTTATACCAAGGTT  82295

Qy   3114  GTTGGCCCAGGGGCGTTAAATTCACAGATTTTTTTAACGAGAAAAACACAGAGAGAGCT  3173
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82296  GTTGGCCCAGGGGCGTTAAATTTATCCGATTTTATAAATGAGAAAAACACAGAGAAAAGCT  82355

Qy   3174  ACCTCAGGTG-TTTTTACCTCAGCACCTTGCTCCTTGTGTTTCCTTGAGATTTTGTAAA  3232
           ||||||||||  |||||||||||||||||||||||||||||||||||||||||||||||||
Db  82356  ACCTCAGGTGTTTTTGCCTCAGCACCTTGCTCCTTGTGTTTCCTTAGAGATTTTGTAAA  82415

Qy   3233  GCTGATAGTTGGAGCATTTTTTATTTTTAATAAAAATGAGTTGGAAAAAAAATAAGA  3292
           ||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db  82416  ACTGATAGTTGGAGCA---TTTATTTTTCTTAATAAAAATGAGTTGGAAAAAAAATAAGA  82471

Qy   3293  TATCAACTGCCAGCCTGGAGAAGGTGACAGTCCAAGTGCAACAGCTGTTCGAATTGT  3352
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82472  TATCAACAACCAGCCTGGAGAAGGTGAGAGTCCAATTGTCAACAACTGTTCGAATTGT  82531

Qy   3353  CTTCCGCTAGCCAAGAACCNATATGGCCTTCTTTTGACAAACCTTGAAAATGTTTATT  3411
           |||||||||||||||||||  |||||||||||||||||||||||||||||||||||||
Db  82532  CTTCTGCTAGCCAAGAACCTATATGGCCCTTCTTTTTAGACAAACCTTGAAAACATTTATT  82590
```

RESULT 13
AP004290/c

LOCUS   AP004290   113201 bp   DNA   linear   PRI 04-DEC-2001
DEFINITION Homo sapiens genomic DNA, chromosome 8q23, clone: KB261B8.
ACCESSION   AP004290
VERSION   AP004290.2 GI:17298202
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Shimizu,N. and Asakawa,S.
  TITLE   Homo sapiens DNA chromosome 8 SEQUENCE
  JOURNAL   Published Only in Database (2001)
REFERENCE   2 (bases 1 to 113201)
  AUTHORS   Shimizu,N. and Asakawa,S.
  TITLE   Direct Submission
  JOURNAL   Submitted (18-OCT-2001) Nobuyoshi Shimizu, Keio University, School
            of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo
            160-8582, Japan (E-mail:nshimizu@dmb.med.keio.ac.jp,
            Tel:81-3-3351-2370, Fax:81-3-3351-2370)
COMMENT   On Dec 3, 2001 this sequence version replaced gi:16303299.
FEATURES   Location/Qualifiers
     source   1..113201
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8q23"
                     /clone="KB261B8"
                     /cell_line="FLEB 14 - 14"
                     /cell_type="pre-pro-B cell"
                     /clone_lib="Keio BAC library"
     repeat_region   131..298
                     /evidence=not_experimental
                     /rpt_family="CT-rich"
     repeat_region   490..512
                     /evidence=not_experimental
                     /rpt_family="AT_rich"

```
Qy   1865  AGTGAGG-CTCCCACAGGCACCGACCAAAACACGGATGAATGAGCCCTTCCAACACCTG  1923
           |||||||  |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  81046  AGTGAGGGCTCCCACAGGCATCAGCAAAACACAGATGAATGAGCCCTTCCAACACCTG  81105

Qy   1924  ACAGAATGAGACCAAACGAGCCCAGCCAGATCGGGAGCAAACCAAGACCATCTGAGGAA  1983
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  81106  ACAAAATGAGACCAAACACAGATCGGGAGCGAAGCAAACCAAGACCATCTGAGGAA  81165

Qy   1984  TGAGAAGTCTGCGAGGCCGGCCGAGGGCCTGAGAACCCCAGGGGCCGA  2043
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db  81166  TGAGAAGTCTGCGAGGCCGGCCGAGGACTCTGCAGAGGCCCTGAGAACTCTGCAGGGGCCAA  81225

Qy   2044  GGAGGGGCGGGGAAGGTCAGCCAGGTTGCCAGCACCACCACGAGCCCGCCTCCCGCCCCC  2103
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  81226  GGAGGCGCG-GGGGAAGGTCAGCCAGGTTGCCAGGAACCTTGCCAGAACCTCCGGGGCCCCTGCCTCCTCCTACCTCC  81284

Qy   2104  CAGGGCTTCTGCGAGCTTCAGCAATCCACTTCACCACTCACTCCTGAACTC  2163
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  81285  CAGGACTTCTGCA-GCTTCAGCAATCCACTTCACCACTCACTCCTTAACTC  81343

Qy   2164  CCACGACGTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAA  2223
           |||||||||||||||||    |||||||||||||||||||||||||||||||||||||
Db  81344  CCACAATGCTGTACCTTTTGGTTGAAGTAACATAGGTGAACGTGTTCAAAGCCATGCAAA  81403

Qy   2224  ATGCACACCCCTTTTCTGTGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAAGGG  2283
           |||||||||||||||||||       |||||||||||||||||||||||||||||||||
Db  81404  ATGCACACCCCTTTTTGTGTGCAAA----TCTCGTACATGTGTGTGCACATATTAGAAGGGG  81459

Qy   2284  AAGATGTTTAAGATATGTGGCCTGGGTTACACAGGGTGCCTGCAGCGGTAATATATTTT  2343
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  81460  ACGATGTAAGATACGTGGCCTGCGGGTTACGCAGGGTGCCTGCAGCGGTAATATGTTTT  81519

Qy   2344  AGAAATAATATCAAATAACTCAACTAACTCAATTTTAATCAATTATTAAATTTTTTT  2403
           ||||||||||||||||||||||||||||||||||||||||||||   ||||||||||
Db  81520  AGAAAACATATCAAATAACTCAACTAACTCAATTTTAATCAATTAAACTAACTATTAATTTTTTT  81579

Qy   2404  TTCTTTTTAAAGAAGAAAGACAGGCTTTCTGAGACTTTAAAGAATAAAGTCTTGGGAGGTC  2463
           |||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db  81580  TTCTTTTTAAAGAAGAAAGACAGGCTTTTCTCAGACTTTAAAGTCT---TTCTTTGGGAGGT  81635

Qy   2464  TCACGGTGTGAGAGGAGTTGAGGCACCGCACCCGACAAAATTCACCCAGGAGGGAAATCTC  2523
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  81636  TCACGGTGTGAGAGGAGTTCGAGGTGACCTACACAAAATTCACCCAGGAGGGAAATCTT  81695

Qy   2524  GTCGGAAGGACACTCACGGCCAGTTCTCGGATCACCTGTGTATGTCAACAGGAGGATACCG  2583
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  81696  GTCGGAAGGACACTCACAGCCAGCAATTCACCTCTGTATCTCAACAGGAGGATACCA  81755

Qy   2584  TCTCCCTGAAGAGGAAACTCTGTGCACTCCCTGTCCTGAGCTCATACACCCATTTCT  2643
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  81756  TCTCCTTGAAAAGGAAACTCTGTCACTCCTGTCCTGAGCTCTATACATCCATTTCT  81815

Qy   2644  CTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGGTCTATATAATTCTC--TGTCTC  2701
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||  |||||
Db  81816  CTTTGCTTCACATGGTTTTTTTAAACTGGTTTTTTGCATACTGGTCTATATAATAATTCTCTATCTC  81875

Qy   2702  TCTCTGTTTATCTCTCCCCCCTCTCCCCCTTCTCCCATCTCCATTCTTTTTGAAT  2761
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  81876  TCTCTGTTTATCTCTCCCCCCTCTCCCCCTACCTTCTCCATTCTCCATTCTTTTTGAAT  81935

Qy   2762  TTCCTACTCCCTTCCATCTCAATCCCGTATCTACGCA------CCCCCCCCCCCAGG  2813
           |||||||||||||||||||||||||||||||||||      |||||||||||||||
Db  81936  TCCCTACTCCCTTCCATCTCAATCCCATATCTACGCAACACCCACCCTCGCCCCCCGCCCCAAG  81995

Qy   2814  CAAAGCAGTGCTGCTTGAGATCCCAACAACAACACAGAGCAAATCACACACACGTCA  2873
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  81996  GAAAGCAGTGCTTGAGATCCCAACACAACAACAGGAACAAAAGCAAAAACACACACGTCA  82055

Qy   2874  GCCTCAACTTACACTTGGTTACTCAAAGAAGAACAAGAGTCAATGGTACTTGCTCCTAGCGTT  2933
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82056  GCCTCAACTTACGCTTGGTTACTCAAAGAAGAACAAGAGTCAATGGTACTTGCTCCTAGCATT  82115
```

```
repeat_region   1347..1650
                /evidence=not experimental
                /rpt_family="AluSc"
repeat_region   3263..3388
                /evidence=not experimental
                /rpt_family="MIR"
repeat_region   3401..3792
                /evidence=not experimental
                /rpt_family="MLT1J"
repeat_region   3796..4090
                /evidence=not experimental
                /rpt_family="AluJb"
repeat_region   4165..4258
                /evidence=not experimental
                /rpt_family="MIR"
repeat_region   complement(4257..4805)
                /evidence=not experimental
                /rpt_family="L2"
repeat_region   4865..5168
                /evidence=not experimental
                /rpt_family="LiMD2"
repeat_region   5178..5495
                /evidence=not experimental
                /rpt_family="L1MC/D"
repeat_region   complement(5495..6055)
                /evidence=not experimental
                /rpt_family="LiMD2"
repeat_region   complement(6057..6348)
                /evidence=not experimental
                /rpt_family="L1M4"
repeat_region   complement(6475..6762)
                /evidence=not experimental
                /rpt_family="AluJo"
repeat_region   complement(6800..7005)
                /evidence=not experimental
                /rpt_family="MLT1J"
repeat_region   complement(7014..7295)
                /evidence=not experimental
                /rpt_family="AluSx"
repeat_region   complement(7402..7456)
                /evidence=not experimental
                /rpt_family="LiMD2"
repeat_region   7690..7746
                /evidence=not experimental
                /rpt_family="AT_rich"
repeat_region   7993..8287
                /evidence=not experimental
                /rpt_family="AluJo"
repeat_region   8341..8400
                /evidence=not experimental
                /rpt_family="AT_rich"
repeat_region   complement(9219..9514)
                /evidence=not experimental
                /rpt_family="AluSp"
repeat_region   9566..9901
                /evidence=not experimental
                /rpt_family="AluSx"
repeat_region   complement(10118..11352)
                /evidence=not experimental
                /rpt_family="MER1B"
repeat_region   complement(11355..11428)
                /evidence=not experimental
                /rpt_family="LiMC1"
repeat_region   complement(11429..11730)
                /evidence=not experimental
                /rpt_family="LiMEc"
repeat_region   complement(11731..12789)
                /evidence=not experimental
                /rpt_family="AluSp"
repeat_region   12790..13085
                /evidence=not experimental
                /rpt_family="LiMEc"
repeat_region   complement(13086..14349)
                /evidence=not experimental
                /rpt_family="AluJb"

                /evidence=not experimental
                /rpt_family="LiMEc"
repeat_region   14367..14502
                /evidence=not experimental
                /rpt_family="L1ME"
repeat_region   14724..14905
                /evidence=not experimental
                /rpt_family="LiME"
repeat_region   15003..15180
                /evidence=not experimental
                /rpt_family="L1M4"
repeat_region   15217..15433
                /evidence=not experimental
                /rpt_family="LiME3"
repeat_region   16753..16837
                /evidence=not experimental
                /rpt_family="AT_rich"
repeat_region   17096..17145
                /evidence=not experimental
                /rpt_family="Tigger1"
repeat_region   17146..17170
                /evidence=not experimental
                /rpt_family="(TG)n"
repeat_region   17171..18147
                /evidence=not experimental
                /rpt_family="Tigger1"
repeat_region   18148..18427
                /evidence=not experimental
                /rpt_family="AluSg"
repeat_region   18428..19574
                /evidence=not experimental
                /rpt_family="Tigger1"
repeat_region   complement(19575..19852)
                /evidence=not experimental
                /rpt_family="AluSg"
repeat_region   19853..20035
                /evidence=not experimental
                /rpt_family="Tigger1"
repeat_region   20272..20565
                /evidence=not experimental
                /rpt_family="AluSq"
repeat_region   complement(20853..20937)
                /evidence=not experimental
                /rpt_family="MLT1H"
repeat_region   complement(21626..21927)
                /evidence=not experimental
                /rpt_family="AluSx"
repeat_region   complement(22312..24467)
                /evidence=not experimental
                /rpt_family="pTR5"
repeat_region   complement(24463..24524)
                /evidence=not experimental
                /rpt_family="HERV9"
repeat_region   complement(24523..24823)
                /evidence=not experimental
                /rpt_family="Harlequin"
repeat_region   complement(24954..25233)
                /evidence=not experimental
                /rpt_family="L1MA8"
repeat_region   complement(25679..25860)
                /evidence=not experimental
                /rpt_family="MER58D"
repeat_region   complement(25861..26177)
                /evidence=not experimental
                /rpt_family="AluSg"
repeat_region   complement(26178..26364)
                /evidence=not experimental
                /rpt_family="MER58D"
repeat_region   26399..26798
                /evidence=not experimental
                /rpt_family="L1"
repeat_region   27355..27395
                /evidence=not experimental
```

```
repeat_region    27958..27979    /rpt_family="AT_rich"
                                 /evidence=not_experimental
repeat_region    29533..29562    /rpt_family="AT_rich"
                                 /evidence=not_experimental
repeat_region    30365..30397    /rpt_family="AT_rich"
                                 /evidence=not_experimental
repeat_region    30655..30712    /rpt_family="AT_rich"
                                 /evidence=not_experimental
repeat_region    30881..30935    /rpt_family="AT_rich"
                                 /evidence=not_experimental

Query Match         64.1%;  Score 2188;  DB 9;  Length 113201;
Best Local Similarity   85.8%;  Pred. No. 0;
Matches 2726;  Conservative 0;  Mismatches 211;  Indels 242;  Gaps 17;
```

```
Qy    1026 CAACCCGGAAAGAACCATCACTGTGAAGGACCACTGTGAAGGCCTGTGCCAGTGCTGAGAT 1085
           ||||||||||||||||||||||||||||||         |||   || |||||||||||||
Db   93577 CAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTGGAGGCTATGCTGGCTGAGAT     93518

Qy    1086 AGAGATTATGAGAGAGCTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTGAAACAACA  1145
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   93517 GGAGATTATGAGAGAGCTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTAA-----   93464

Qy    1146 AGCCAATCTGATCCCAGGTGAACCTTCAGCGCACTTGGCATCTTTTCAAGAGGACTGTC   1205
Db   93464 ------                                                       93463

Qy    1206 CGTGCTATCTCCACCAGCAGGGCCCGGGAGCTCCCCCGCTGCCCCCTACCACCCCCTT    1265
Db   93463 ------                                                       93463

Qy    1266 CACTACCCACTCCGGATAC-TTCTCAGCTGTACCCCCATCACCAGTTGCCCGTTCC     1324
           |||||||||||| |||||| ||||| ||||||||||||||||||||||||||||||
Db   93463 ---TACCCACTCCAGATACTTTCTCCAGCTGTACCCCCATCACCAGTTGCCCGTTCC     93407

Qy    1325 CGCATCATCACTCTTATCCAGCAGCAGAGATTGTGAATCTTCATCCCAACCCAGGCTG    1384
           ||||||||||| ||||| ||||||||||||||| |||||||| |||||||||||||||
Db   93406 CGCATCAGCATTCTTATCCAGCAGCAGAGATTGTGAATCTTCATCCCAACCCAGGCTG    93347

Qy    1385 TGGGCGCCATCAGTTGAGAATTCAGATTCCTCCTCACCTGACAGTGGCGAGATTCGCCGGAG 1444
           ||||||||||||||| |||||||||||||||| |||||||||||||||||||||||||||||
Db   93346 TGGGTACCATCATCGTGGGAAGAAATTCAGATTCCTCACCTGCGCCAGAG            93287

Qy    1445 CCTCTATCAAGATTGCCCCTGCGGAAGCCCGAAGCGTCAGCGAAAGGATGGTCATCATCA 1504
           |||||||||||
Db   93286 CCTCCA-                                                        93281

Qy    1505 CCCGGCCACCGGAGCCCAGTTCAAGGCCCAGGACGATCTTTTGGGAAACTGAAAGAGG    1564
           |||||||||||||||||||| ||||||||||||||||||||||| ||||||||||||
Db   93280 --------TCAAGGCCCAGGACGATCTTCTCGGGAAACTAAAAGAGAG            93242

Qy    1565 AAAACTTCTTTAACCCAAAGAAGAAGTCAAGAGAGAGACCTGGAAGCGCATATCAGAGTGCCCTCTT 1624
           ||||||||||| |||||||||||||||
Db   93241 AAAACTTCTTTAACCCAAAGAAGAAGTGAAGCGTGGAGCCCATATCAGAGTGTCTCTCT      93182

Qy    1625 CCACAGCTGCGCGGTGATTGCCAAGACCTGGCAAGACGTCAGCGAAGGATGCGAACTAA  1684
           ||||||||||||||||||| ||||||||||||||||||||||||||
Db   93181 CCACAGCTGGCGGTGATTGGCAAGACTGTGATTGGCAAGACTGCATAACTTAA          93122

Qy    1685 CCAGTGCAGAGTCATCGTTGTCAGATGAAAATGCCCAGATATCAGGAAGTGATCG      1744
           ||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   93121 CTAGTGCAGAAATCATTGTGCCGTCGTGACCCCCATATCAGGAAGTGATCA           93062

Qy    1745 TCAGAATTATCGGGCACTTCTTTTGCTAGCCAGCCGCAAGATCAGGGAAATTG         1804
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   93061 TCAGAATTATCGGGCACTTCTTTTGCTAGCCAGCCGCAGTATCAGGGAAACTG         93002

Qy    1805 TACAACAGGTGAAGCAGCAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAGCA    1864
           ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   93001 TACAAGAGGTGAAGCAGCAGCAGAGAATACCCTCAGGGAGTCGCCTCACAGCGCAGCA    92942

Qy    1865 AGTGAGG-CTCCCACAGCGATCAGCAGCCAGCAAAACAACAGGATGAATGTAGCCCTTCCAACACCTG 1923
           ||||||||
Db   92941 AGTGAGGAGTGAGCCCTTCCAACAGATCAGCAGCAAAACAACAGGCCCTTCCAACACCTG   92882

Qy    1924 ACAGAATGACAACACGCAGCCAGCCAGCAAAGACCATCTGAGAA                1983
           ||||||||||||||||||||||||||||||||||||||||||||
Db   92881 ACAAATGACAACACAGCCAGCCAGCAAAGACCATCTGAGAA                   92822

Qy    1984 TGAGAAGTCTGCGGAGGCGGCGGAAGGGCGGCGCTCTCGCAGGGGCCGA           2043
           |||||||||||||||||||||||||| |||||||||||||||||||||
Db   92821 TGAGAAGTCTGCGGAGGCGGCGGAATCCCAGGGGCCAA                      92762

Qy    2044 GGAGGGGCGGGGAAGGTCAGCGAAGGTTGCCCAGCATATA                    2103
           |||||||||||||||||||||||||||||||||||||||
Db   92761 GGAGGCG-GGGGAAGGTCAGCGAAGGTTGCCCAGCATATG                    92703

Qy    2104 CAGGGCGTTCTGCAGGCTTCAGCCATCCACTTCACCATCCACTCACTGCTCGGATCTCCTGAACTC 2163
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   93637 AAATGAACATGAAACAGGGACCAAGATAACATCTCATCTTTCATCCTTGCAGGATTTGAGCATATG 93578
```

```
Qy     247 GGTAAGTGAAATGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTA   306
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   94347 GGTAAGTGAAATGCATGGGAAAATCATGGAAGTTGATTACTCGGTCTCTAAAAAGCTA   94288

Qy     307 AGGAGCAGGAAAATTCAGATTCGAAACATCCCTCCTCACCTGACAGGTGTTGGAT     366
           ||||||||||||||| ||||||||||||||||||||||||||
Db   94287 AGGAACAGGAAAATTCAGATTCGAAACATACCTCCTCACCTGCGGCGATGTTGGAT    94228

Qy     367 GGACTTTGGCTCAATATGGGACAGTGGAGAATGTCAACAAGTCAACACAGACACAGAA 426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   94227 GCACCTTTGGCTCAACAGCGGGATAGTGGAGAATGTGGAGAACAAGTCAACACAGACAGAA 94168

Qy     427 ACCGGCCGTTGTCAAGCTCACATATGCATATTGAAGCCATGAGAAG              486
           |||||||||||||||||||||||||||||||||||||||
Db   94167 ACTGCTGTGTCAATGCTCACATATGCATATGCAAGCCATGAGGAG               94109

Qy     487 CTAAGGCGGCATCAGTTGAGAATTCCTACATCCCGGATGAAGAG                546
           ||||||||||||||||||||||||||||||||
Db   94108 CTAAGGCGGCATCAGTTGAGAACTATTCCTACATCCTGGATGAAGAG             94049

Qy     547 GTGAGCT-CCCCTCGCCCCCAGCGTGGGGACCACTCTTCCCGGAGCA             605
           |||||||  ||||||||||||||||||||||||||||||
Db   94048 GTGAGCTCCCCTTACCCCCTCAGCGACCAGCCCAGGGTGGGGACCACTCTT--CCCGGAGCA 93990

Qy     606 AGGCCAGCGCCCCTGGGGCACTTCTCAGCCAGACAGATTGATTTCCCGCTGCGGATCCT  665
           |||||||||||||||||||||||||||||||||||||||||||||
Db   93989 AGGTCACG--CCTGGGGCGTCTTCTCAGCCAGACAGACTGATTTCCCGCTGCGGATCCA   93931

Qy     666 GGTCCCCACCCAGTTGTTGTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAGAACAT 725
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   93930 -------ACCCAGTTGTTGTGGTGCCATGGTCTGAAAGGAGGGCTTGAACATAGTGAACAT 93878

Qy     726 CACTAGCAGGACCCGGTAGATATCCATAGAAAAGAACTCTGAGCTGCAGA          785
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db   93877 CACTGAGCAGCAGACCCAGTCCCGGTAGACATCCATAGAAAAGAACTCTGAGCTGCAGA   93818

Qy     786 GAAGCCTGTCACCATCCATGCCACCCACCCCAGAGGGGGACTTCTGAAGCATGCGCCACACGATTCT 845
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   93817 GAAGCTTGTCACCATCCATGCCACCCACCCCAGAGGGGGACTTCTGAAGCATGCGCCACACGATTCT 93758

Qy     846 TGAAATCATGCGAGAAAGAGGCAGCAGATGAGAAACTAGCCGAAGAACTTCCTCTGAAAAT 905
           ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   93757 TGAAAGCATGCGAAGAGAAAGAGGCAGCAGATGAGAAACTAGCCGAAGAGATTCCTCTGAAAAT 93698

Qy     906 CTTGGCACACAATGGCTTGTTGGAAGACTCATCTTCATCCTTGCAGAAATTTGAAGAA     965
           |||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||||
Db   93697 CCTTGGCCCACAATGGCTTGTTGTTGGAAGACTGATTGGAAAGAAGGCAGAAATGTGAAGAA 93638

Qy     966 AATTGAACATGAAACAGGGACCAAGATAACATCTCATCTTTCATCCTTGCAGGATTTGAGCATATA 1025
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db   93637 AAATGAACATGAAACAGGGACCAAGATAACATCTCATCTTTCATCCTTGCAGGATTTGAGCATATG 93578
```

```
Db    91631 ACTTCAGGTGTGTTTTTGCTCAGCACCTTGCTCTTGTGTTTCTTAGAGATTTGTGTAAA 91572
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     3233 GCTGATAGTTGGAGCATTTTTTATTTTTTAATTTTTTAATAAAATGAGTTGGAAAAAAATAAGA 3292
            |||||||||||||||||||||||||||         ||||||||||||||||||||||||||
Db    91571 ACTGATAGTTGGAGCA----TTTATTTCTTAATAAAATGAGTTGGAAAAAAATAAGA 91516
Qy     3293 TATCAACTGCCAGCCTGGAGAAGGTGACAGTGCAACAGCTGTTCTGAATTGT 3352
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    91515 TATCAACAACCAGCCTGGAGAAGGTGAGAGTGCCAATTGCGCAACAACGTTCTCGAATTGT 91456
Qy     3353 CTTCCGCTAGCCAAGAACCNATATGGCCTTCTTTGGACAAAATGTTATT 3411
            |||||||||||||||       |||||||||||||||||||||||||||
Db    91455 CTTCTGCTAGCCAAGAACCTATATGGCCTTCTTTTGACAAAACCTTGAAAACATTATT 91397
```

RESULT 14
BC023758
LOCUS   BC023758   3557 bp   mRNA   linear   ROD 16-APR-2003
DEFINITION Mus musculus RIKEN cDNA C330012H03 gene, mRNA (cDNA clone
            IMAGE:5354659), partial cds.
ACCESSION   BC023758
VERSION   BC023758.1 GI:23958572
KEYWORDS   .
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 3557)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C. Rodriguez,A.C., Grimwood,J. Schmutz,J. Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2 (bases 1 to 3557)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (05-FEB-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue procurement: Jeffrey Green M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
```

```
Db    92702 CAGGGACTTCTGCA-GCTTCAGCAATCCACTTCACCTCACCCACTCAGATCTCTCCTTAACTC 92644
            ||||  |||||||||                              |||||||  |||||
Qy     2164 CCACGACGCTATCCCTTTAGTGACTAACATAGGTGAACGTGTGTTCAAAGCCAAGCAAA 2223
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    92643 CCACAATGCTACCCTTTTGGTGAGTAACATAGGTGAACGTGTTCAAAGCCATGCAAA 92584
Qy     2224 ATGCACACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAAGGG 2283
            |||||||||||||||||||||        |||||||||||||||||||||||||||||
Db    92583 ATGCACACCCCTTTTTGTGTGGCAAA----TCTCTGTACATGTGTGTGCACATATTAGAAGGGG 92528
Qy     2284 AAGAGTGTAAGATATGTGGCCTGTGGTTACACAGGGTGCCTCAGCCGGTAATATATTTT 2343
            |||||||||||||||||||||||||||||||||||||||||||||||||  |||||||
Db    92527 ACGAGTGTAAGATACGTGGCCTGCGGGGTTACGCAGGGTGCCTGCAGCCGGTAATATGTTTT 92468
Qy     2344 AGAAATAATATATCAACTAACTAACTAATTCCAATTTTTAATCAATTATTAATTTTTTTT 2403
            |||||||   ||  ||||||||||||       ||||||  |||    ||||||||||
Db    92467 AGAAATAACATATCAATCAATCAACTCACTAATTTTAATCAACTATTAATTTTTTTT 92408
Qy     2404 TTCTTTTTAAAGAAGAGGAGGCTTTTCTAGACTTAAAGAATAAAGTCTTTGGGGAGTC 2463
            ||||||||||||||||||||||||||||         ||||   |||||||||||||
Db    92407 TTCTTTTTAAAGAAGAGGAGGCTTTTCTAGACTTTAAAGTCT----TTCTTTGGGGAGT 92352
Qy     2464 TCACCGGTGTGAGAGGACGAGGCTTGAGGCCACCCGCACACAAAATTCACCCGAGGGAAATCTC 2523
            ||||||||||||||||||||||||||||||||  ||| || ||||||||||||||||||
Db    92351 TCACCGGTGTGAGAGGACGAGGCTTGAGGTGACCTACACACAAAATTCACCCGAGGGAAATCTT 92292
Qy     2524 GTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTCAACAGAAGGGGATACCG 2583
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    92291 GTCGGAAGGACACTCACAGCAGTTCTGGATCATCTGTGTATATCAACAGGAGGGGATACCA 92232
Qy     2584 TCTCCTTGAAGAGAAAACTTCTGTCACTCCTCAGCTCATCACACCCATTTCT 2643
            |||||||||||||||||||||  |||||||  ||||||||||||||||||||
Db    92231 TCTCCTCCTCCAATCCATCTGTCCATCCCATCCGTCTAGCTTATCATCCATTTCT 92172
Qy     2644 CTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGCTATATATAATTCTC--TGTCTC 2701
            ||||||||||||||||||||||||||||||||||||||||||| |||||||   |||||
Db    92171 CTTTGCTTCATGGGTTTTAAACTGGTTTTTTCGCATACTGCTATATATTCTCTATCTCTC 92112
Qy     2702 TCTCTGTTATCTCTCCCCCTCCCCCCCCCCTCTTCTCCATTCTTTGAAT 2761
            |||||||||||||||||||      ||||  |||||||||||||||||
Db    92111 TCTCTGTTTATCTCTCCACCTCCCCCCCCCCCACTTCTCCATTCTTTTGAAT 92052
Qy     2762 TTCCTCATCCTCCTCCATCTCAATCCCGTATCTACGGCA------CCCCCCCCCCAGG 2813
            ||||||||||||||||||||||||||                 ||||||||||||||
Db    92051 TCCCTCATCCTCCTCCATCTCAATCCCATATCTACGCAACACCACCCTCGCCCCCGCCCAAG 91992
Qy     2814 CAAAGCAGTGCTCTGAGTATCACATCACACAGAGGAACAAAAGGCAGACACACAAACCA 2873
            |||||||||||||||||||||||  |||| |||||||| |||||||||||||||||
Db    91991 GAAAGCAGTGCTTTGAGTGTCTTTGAGTGTACTGTCCTAGCGTT 91932
Qy     2874 GCCTCAACTTACACTTGGTTACTCAAAGAACAAGAGACAAGAGTACTTGTCCTAGCGTT 2933
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    91931 GCCTCAACTTACACGCTGGTTACTCAAAGAACAAGAGACAAGAGTACTTGTCCTAGCATT 91872
Qy     2934 TTGGAAGAGAAAACAGAGACCACCAACCAACCAACCAAACAAGAAAAAAATTC 2993
            ||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db    91871 TTGGAAGAGAAAAACAGAGAACCAGGAACCAACCAACCAACAAACAAGAAAAAAATTC 91812
Qy     2994 CACAATGAAAAGAAATGTATTTTTGTCTTTTTGCATTTTGGTGTATAAGCCATCAATATTCAG 3053
            ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    91811 CACAATGAAAAGAAATGTATTTTGTCTTTTTGTTGATTTTTTGGTGTATAAGCCATCAATATTCAG 91752
Qy     3054 CAAAATGATTCCTTTCTTTAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTT 3113
            ||||  |||||  |||||||||||||| |  |||      |||||||||||||||
Db    91751 CAAAATTATTTCTTTCTTTTTTTTAAAAAAATGTGAAGGAAAGTAGCAATTTATACAAGGTT 91692
Qy     3114 GTTGGCCAGGGCGTTAAATTCACAGATTTTTTACGACAGAAAAACACAGAAGAGAGCT 3173
            |||||| |||||||||||||||||||||||| |||||||||||||||| |||||||||
Db    91691 GTTGGCCCAGGCGTTAAATTTACCGATTTTAAATTTACCGATTTTAATGAGAAAATGAGAAAAACACAGAAAAAGCT 91632
Qy     3174 ACCTCAGGTG-TTTTTACCTCAGCACCTTGCTCTTGTGTTTCCCTTGATGAGATTTTGTGAAA 3232
            |||||||||| |||||||||||||||||||||||||||||||  ||||||||||||||||||
```

Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 55 Row: o Column: 24
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, GenomeScan gene prediction.

FEATURES   Location/Qualifiers
     source   1..3557
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5354659"
                     /tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating
                     ductal carcinoma, 5 month old virgin mouse."
                     /clone_lib="NCI_CGAP_Mam6"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"

ORIGIN

Query Match   63.5%;   Score 2165;   DB 10;   Length 3557;
Best Local Similarity 83.0%;   Pred. No. 0;
Matches 2820;   Conservative   0;   Mismatches 476;   Indels 100;   Gaps 27;

Qy   58 GGGAGAGACGGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCCGGTCACCGCC 117
         |||||||||||||||||||||||||| ||||| ||||||||||||||||||||||||
Db   23 GGGAGAGACGGGATGATGAACAAGCTGTACATTGGGAACCTGAGCCCCGGTCACCGCC 82

Qy   118 GACGACCTCCGGCAGCTCTTTGGGGACACAGGAAGTGCCCCTGGCGGGGACAGGTTCCTGTG 177
         |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   83 GACGACCTCCGGCAGCTCTTCGGGGACACAGGAAGTGCCCCTGGCGGGGACAGGTCCTACTC 142

Qy   178 AAGTCCGGCACGCCTTCGTGGACTACCCCGGACTACCCCGGACAGGAAACTG 237
         |||||||||||||||||||||||||||||||||||||||||||||
Db   143 AAGTCCGGCTACGCCTTCGTGGACTACCCCGGACTACCCCGGACAGGAAACTG 202

Qy   238 ACCCTTCGGGTAAAGTGCATGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCT 297
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   203 ACCCTTCGGGTAAAGTGCATGGAATTGCATGGGAAAATCATGGAAGTTGACTACTCAGTCT 262

Qy   298 AAAAAGCTAAGGAGCAGGAGAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGGAG 357
         ||||||||||||||||||||||||||||||||||||||||||||||
Db   263 AAAAAGCTAAGGAGCAGGAGAAGATTCAGATTCGGAACATCCCTCCTCACCTGCAGTGGGAG 322

Qy   358 GTGTTCGATGACTTTTGGCTCAATATGGGACAGTGGAAATGTGAAGAAGTGTCAACAGCC 417
         |||||||||||||||||||||||||| || |||||||||||||||||||
Db   323 GTGTTCGATGGGCTGTTGGCTCAATATGGGACAGTGGAAATGTGAAGAAGTGTCAACAGCC 382

Qy   418 GACACAGAGAAACCGCCGTTGTCAACTGCCCGTTGTCAACTCGGACAGAGAAATAGCC 477
         |||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   383 GATACAGAGAAACTGCCGTTGTCAACTGCCCGTTGTCAACTCGGACAGAGAGAAGCTAGCT 442

Qy   478 ATGGAGAGAGCTAAGCGGGCACGGGGCCACCGCCCTGGGGCACTTCTCAGCCAGACAGATTCCCGCTG 537
         ||||||||||||||||||||||||||||||||||||||||||||||
Db   443 ATTGAGAGAGCTAAGCGGGCTCGTGTTGGCTGGGCATCAGTTTGAGAGACTACTCCTTCAAGATTTCCTACATCCCG 502

Qy   538 GATGAAGAGGTGAGCTCCCCTTCGCCCCGAGCCCAGCGTGGGGACCACTCTTCC 597
         ||||||||||||||||||||||||||||||||||||||||||
Db   503 GATGAAGAGGTGAGCTCCCCTTCGCCCCGAGCCCAGCGTGGGGACCACTCATCG 541

Qy   598 CGGGAGCAAGGCCACGCCCCTGGGGCACTTCTCAGCCAGACAGATTGATTTCCCGCTG 657
         |||||||||||||||||||||||||||||||||||||||||||||||
Db   542 CGGGAACAAGGCCACGCCCCGGGAGCTCTTCTCAGCCAGACAGATTGATTTCCCGCTG 601

Qy   658 CGGGATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCCGGAAAGGAGGGCTTGACCATA 717
         ||||||||||||||||||||||||||||||||||||||||||||||
Db   602 CGGGATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCCGGAAAGGAGGGCTTGACCATA 661

Qy   718 AAGAACTCACTAAGCAGACGACCCAGTCCCGGTCCGGTAGTATCCATAGAAAAGAAGAACTCTGA 777
         |||||||||||||||||||||||||||||||| |||||||||||||||||||
Db   662 AAGAACATCACTAAGCAGACGACCCAGTCCCGGTAGACATCCACAGAAGAAGAGAACTCTGGG 721

Qy   778 GCTGCAGAGAAGCCTGTCACCATCCATG|||||CCACCCCAGAGGGGACTTCTGAAGCATGCCGC 837
         |||||||||||||||||||||||||||||||||||||||||||||
Db   722 GCTGCAGAGAAGCCTGTCACATCCATGTACCCCAGAGGGGACATCTGAAGCATGCCGC 781

Qy   838 ATGATTCTTGAAATCATGCAGAAAGAGCAGATGAGAGCGAAACTAGCCGAAGAGATTCCT 897
         ||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   782 ATGATTCTTGAGATTATGCAGCAAAAAGAAGCTGCAAAAACTGGCTGAGAGGTTCCT 841

Qy   898 CTGAAAATCTTGGCACACAATGGCTTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAAT 957
         ||||||||||||||||||||||||||||||||||||||||||||||
Db   842 CTGAAAATCCTGGCCCCCAATGGCTTCGTTGGAAGACTGATTGGCAAAGAAGGCAGAAAC 901

Qy   958 TTGAAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTCGAGGATTTG 1017
         ||||||||||||||||||||||||||||||||||||||||||||||
Db   902 CTGAAGAAAAATGAACATGAGACATGAGGGACCAAGATAACCATCTCATCCTTGCAGGATTTG 961

Qy 1018 AGCATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGGCTGTGCCAGT 1077
         |||||||||||||||||||||||||||||||||||||||||||||
Db   962 AGCATTTATAACCCCGAGAACATCACCGTGAGGGGCACCATTGAGCCCTGTGCCAAT 1021

Qy 1078 GCTGAGATAGAGATTATGAAGAAGCGTGCGTGAGGAGCCTTGAAAATGATAATGCTGGCGTGTT 1137
         ||||||||||||||||||||||||||||||||||||||||||||||
Db 1022 GCTGAGATAGAGATTATGAAGAAGCGTCCGAGAGCCTTTGAGAACGACATGCTGGCCATCTTTCAACA 1081

Qy 1138 AACCAACAAGCCAATCGATCGTGATCGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACT 1197
         |||||||||||||||||||||||||||||||||||||||||||||
Db 1082 AACCAACAAGCCAATCGATCTGATCCGGGTAAAACCTCAGTGACTTGGCATCTTTTCGACT 1141

Qy 1198 GGACTGTCCGTGTATCTCCACCAGCACGAGCTGCCCCTGGGCGGGACAGGTTCCTGTG 1257
         ||||||||||||||||||||||||||||||||||||||||||||||
Db 1142 GGACTGTCTGTGCTTCCTCGCCACCACCTCCGACGCTGTACCCCCAGTTCCTCCCTAT 1201

Qy 1258 CACCCCTTCACTACCACCTCCGGATACTTCTCCGACGCCTGTACCCTCCATCCACCAGTTTGC 1317
         |||||||||||||||||||||||||||||||||||||||||||||
Db 1202 CACCCCTTTGCTACCACCTCCGGATACTTCTCCGACGCCTGTACCCTGCACCATTTCGGC 1261

Qy 1318 CCGTTCCCGCATCATCATCACTCTTATCCGAGGAGAGATTGTGAATCTTCATCCCACC 1377
         ||||||||||||||||||||||||||||||||||||||||||||
Db 1262 CCATTCCCACATCATCATCACTACTCCTACCCCAGCAGGAGACTGTAAGTCTCTCTTCATCCCAACC 1321

Qy 1378 CAGGCTGTCGGGCGCCATCATCGGGAAGAAGGGCACACATCAAAACAGCTGGCGAGATTC 1437
         |||||||||||||||||||||||||||||||||||||||||||||
Db 1322 CAGGCTGTGTGGGTGCTATCATCGGGAAGAAGGGCACACACACAAACAGCTCGCTGCGATTT 1381

Qy 1438 GCCCGGAGCCCTATCAAGATTGCCCCTGCGAAGGCAACATCCCAGAGCGTCAGCGAAAGATGGTC 1497
         |||||||||||||||||||||||||||||||||||||||||||||
Db 1382 GCTCGTGTGCCTTCCATCAAGATTGCTCCAGCAAGTTCAGGAGCCGATCTTTGGGAAACTG 1441

Qy 1498 ATCATCACCCGGCACCCGAAGCCCAGTTCAAGGCTCAGGACGCATATCAGAGTG 1557
         |||||||||||||||||||||||||||||||||||||||||||
Db 1442 ATCATCACTGTCTCCTCGAAGCCCAGTTTAAGGCTCAGGACCAGTTTGGGAAACTG 1501

Qy 1558 AAAGAGGAAAACTTCTTAACCCCAAAGAAGAAGTGAAGCTGGAAGCGCATATCAGAGTG 1617
         |||||||||||||||||||||||||||||||||||||||||||||
Db 1502 AAGGAAGAAAAACTTCTTTATCCAAAGAAGAGAAGTGAAGCTGGAAGCCCACATCCGAGTC 1561

Qy 1618 CCCTCTTCCACAGCTGGCCGGGTGATTGGCGGCCCAGCGTGGGAAAGGTGGCAAGACCGTGACGAACTGCAG 1677
         ||||||||||||||||||||||||||||||||||||||||||||
Db 1562 CCCTCGTCGACCGGGTGGGCCCCGGGTGATTGGCGGCCCAGCAAACCGTGACGAGCTGCAG 1621

Qy 1678 AACTTAACAGTGCAGAAGTCATCGTGCCTGCTTCGTGCCTCGTGACAGCCAGAATGAGGAA 1737
         ||||||||||||||||||||||||||||||||||||||||||||||
Db 1622 AACTTGACAAGTGCAGAAGTGTCGTCGTGCCTCGTGACAGCCAGAGAATGAGGAA 1681

Qy 1738 GTGATCGTCAGAATTATCGGGACACTCTTCTTCGTCTAGCCAGCGCAGATCAGG 1797
         ||||||||||||||||||||||||||||||||||||||||||||||
Db 1682 GTGATCGTCAGAATTATCGGGCATTTTTTGCTAGCCAGACTGCACAACGCAAGATCAGG 1741

Qy 1798 GAAATTGTACAACAGTGAAGCAGCAGAGAAATACCCTCAGGGAGTCGCCTCACAG 1857

```
Db  1742  GAAATTGTACAGCAGGTGAAGCAGTGAGGGAGCAGGAGCAGAGATACCCTCAGGGAGTCGCCCCACAG  1801
Qy  1858  CGGCAGCAAGCAGTGAGGCTCCCACAGCACCACGATGAATGTAGCCCTTCAA  1917
Db  1802  CGCAGCGAGTGAGGCTCCCACAG--CACCAGCAAGCAACCGATGAATGTAGCCCTCCCA  1859
Qy  1918  CACCTGACAGAATGAGACCAAAGTCGCGGAGCCAGACGGAGCAAACCAAAGACCAT-C  1976
Db  1860  CACCTGACAG-ATGAGACCAAAC---AGCCAGCAGATCGGAAGCAAACCAAAGAGCATCC  1915
Qy  1977  TGAGGAATGAGAAGTCGCGGAGCGCCCAGGGACTCTGCCGAGGCCCTGAGAACCCCAG  2036
Db  1916  CGAGGAGTGCCAGTCTGCGACAGCAGCAGCCAGGGCCTGC-----AGACCTCTACACATCCTG  1970
Qy  2037  GGGCCGAGGAGGGCGGGGGAAGGTCAGCCAGGTTGCCAGAACCACCG--AGCCCGGCT  2094
Db  1971  GGATCCAGGAGGGGCACAGGGGAAG--GCCCAGGTGTCGCCAGAAACACCCGCTTGGGCTGCCC  2027
Qy  2095  CCCGCCCCCAGGGGCTTCTGCGAGGCTT--CAGCACCATCCACTTCCACCATCGGGATCT  2152
Db  2028  CCAGCTTCCCCTGGCTTCTGCGAGGCGATAACAGCAGCCCATCCACCAACTCAGAT-T  2086
Qy  2153  CTCCTGAACTCCCACGACGCGTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAA  2212
Db  2087  CTCCTCAGTTCCCAGACGCGTATCCCTTTCGGTTGAACTAACATAGGTGAACATGCTCAA  2146
Qy  2213  AGCCAAGCAAAAATGCACACCCTTTTC----TGTGGCAAATGTCTCTGTACATGTGTGT  2268
Db  2147  AGCAAAGCAAAAATTCCTAGCCGTTTTCTTGTTCTTGGAAGTTGTCTCTGTACATGTATGT  2206
Qy  2269  ACATATTAGAAAGGAA-GATGTTAAGATATGTGCCCTGTGGGTTACACAGGGTGCCTGC  2327
Db  2207  ACATATCAGAAGGGAAGGATGGATGTTAAGAATGTGCCCTGTGGGTTACACAGGGTGCCTGC  2266
Qy  2328  AGCCGTAATATATTTTAGAAATATATATCAAATAACTCCAATTTTAAATC  2387
Db  2267  AGCGTTAATATTTTTAGAAATATATATCAAATAACTCCACTAACAACAATTTTACTC  2326
Qy  2388  AATTATTATTTTTTTTTCTTTTAAAGAAAGCAAGCAGGCTTTTCTAGACTTTAAAGAATA  2447
Db  2386  AGTGATTAATTTTTGTTTTTCTTTTTTGAAAGAGAAGCAGGCTTTTCTAGACTTCCAAGAATA  2386
Qy  2448  AAGTCTTTGG-GAGGGTCTCACGGTGTAGACAGAGACTTTGAGGC-CACCCGCACAAAATT  2505
Db  2387  AAGTATTGGGTGAGGGTCCACTGGTATAGAAAGGAACTTTGAGCCACCACCGCACAAAAATT  2446
Qy  2506  CACCCAGGAGGGGAAATCTCGTCGGAAGGACACTCACGGACAGTTCTGGATCACCTGTGTATG  2565
Db  2447  CACTTGGAGGGGGAAATCTTGTC-AAGGACACTTAATGCAGTTGTGGGTCACCTGTGTCTG  2505
Qy  2566  TCAACAGAGGGATACCGTCTCCTTGAGAGAGAAAACTCGTGTCACTCCTCCATGCCTGTCTA  2625
Db  2506  TCAACAGAAGGGATACCG--TCCCTGACGAGGGGCCCTAGGCCCTGGTCACCTGTTTCTC  2563
Qy  2626  GCTCATCACACCCATTCTCTT----------TGCTTCACAGGTTTTAA  2663
Db  2564  CCACACACCCATTGTCCAGTTTCACAGGTTTTTTAAACTGATGATTTTTGGTTTGTTTTGT  2623
Qy  2664  ACTGGTTTTTTTGCATACTGCTATATAATTCCTGTCTTCTGTTTATCTCTGCCTCC  2723
Db  2624  TTTTGTTTTTTTGCATATCGTATATAGTTGTCTGTCTCTC------CTTATCTCCCTCCT  2677
Qy  2724  CTCCCCTCCCCCTTTCTTCTCCATCTCCATTCTTTTTTGAATTCCTCATCCTCCATCTCAAT  2783
Db  2678  AACCTTCCTCTGTCTTCTTTACCCAAAATTCTTTTTGAGTCCTTCGTTTCTTTTCTTGCTC  2737
Qy  2784  CCCGTATCTCTACGCACC---CCCCCCCCCAGGCAAAGCAGTGCTTCTGAGTATCACATCA  2840
Db  2738  CCCGTATCTCTATGCACCCCTCCCCCCTCCCCCCCAGGCAAAGCAGTGCTCTGAATATCATACCA  2797
Qy  2841  CACAAAGGAACAAAAGCGAAACACACAAACCAGCCTCAACTTACACTTGGTTACTCAAA  2900
```

```
Db  2798  CACACAAGGAACAAATGCGAACCACACAGACCACCAGCCTCAGCTTACACTCGGTTACTCAAG  2857
Qy  2901  AGAACAAGAGAGTCAATGGTACTTGTCCTCCTACCGTTTTGGAAGAGAAAAACAGGACCCACCA  2960
Db  2858  CGAACAAGAGTCGGCGGTACTTGTCCCAGC-ATTCGGAAGAGAAGCAGGATCCCAGCA  2916
Qy  2961  AACCAACCAATCACCAACCAAAGAAAAATTCCACAATGAAAGAATGTATTTGTCTT  3020
Db  2917  AATAAACCAGCCAACCAAACAAAG-AAAAATCCCAACAAAGAAGAATGGA---CTCTCT  2971
Qy  3021  TTGCATTTTGGTGTATAAGCCATCAATATTCAGCAAAATGATCCTTTCTTTAAAAAAA  3080
Db  2972  CTCCATCTTGGTATACAAGCCATCAACATTTAGTGAAATTCTTTCTTTCTTTAAAAATTT  3031
Qy  3081  AAATGTGGAGGAAATGAGAAATTTACCAAGGTGTTGTTGCCCAGGGCGTTAAATTCACAGA  3140
Db  3032  AAAGTGGAGGAAAATAGCAATTTATACGAGGTTGTTGGCCCAGGGCGTTAAATTTACAGA  3091
Qy  3141  TTTTTTAACGAGAAAAACACACAGA---AGAAGCTACCTCAGGTG-TTTTTACCTTCAGC  3196
Db  3092  TTTTTTTTAAATGAGAAAAACACACAAATAAAAGCTACCTCAGGTGTTTTTTACCTTCAGC  3151
Qy  3197  ACCTTGCTTCCTTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGGAGCATTTTTTA  3256
Db  3152  ACCTTGCTTCCTTGTTTCCCTTAGAGATTTTGTAAAACTGATAGTTGGAGCA-TTTTTTA  3210
Qy  3257  TTTTTTAATAAAAATGAGTTGGAAAAAAAAAGATATCAACTGCCAGCCTGGGAGAAGG  3316
Db  3211  TTTTTTAATAAAAATGAGTTGAAAAAAAAAAAAG--ATAGCTGCCAGCCCCAAGCAGG  3268
Qy  3317  TGACAGTCCAAGTGTCAAAGCTGTTCTGAATTGTCTTCCGCTAGCCAAGAACCNATAT  3376
Db  3269  CGCAGCCCAGGTGTCCGCAGCTCTTC-----TGTCTTCCGCTAGCCAAGAGCCTACGT  3323
Qy  3377  GGGCCTTCTTTTGGACAAACCTTGAAAAATGTTTATTT  3412
Db  3324  GGCCTTCTTGTGGACAAAACCTAAAAATGTTTATTTT  3359
```

```
CDS          /gene="IMP-2"
             /note="alternative start-codon gives an additional
             N-terminal methionine"
             50. .1846
             /gene="IMP-2"
             /note="binds multiple sites in IGF-II mRNA 5'UTR"
             /codon_start=1
             /product="IGF-II mRNA-binding protein 2"
             /protein_id="AAD09827.1"
             /db_xref="GI:4191610"
             /translation="MNKLYIGNLSPAVTADDLRQLFGDRKLPLAGQVLLKSGYAFVDY
             PDQNWAIRAIETLSGKVELHGKIMEVDYSVSKKLRSRKIQIRNIPPHLQWEVLDGLLA
             QYGVVENVBQVNTDTETAVVNVYATREAKIAMEKLSGHQFENYSFKISYIPDEEVS
             SPSPPQRAQRGDHSSREQGHAPGGTSQARQIDFPLRIIVPTQPVGAIIGKEGLTIKNI
             TKQTQSRVDIHRKENSGAAEKPVTIHATPEGTSEACRMLEIMQKEADETKLAEEIPL
             KILAHNGLVGRLIGKEGRNLKKIEHETGTKIITISSLQDLSIYNPERTITVKGTVEACA
             SAEIBIMKKLREAFENDMLAVNQQANLIPGLNLSALGIPSTGLSVLSPPAGPRGAPPA
             APYHPFTTHSGYFSSLYPHHQFGPFPHHHSYPPQEIVNLFIPTQAVGAIIGKKGAHIK
             QLARPAGASIKIAPAEGPDVSERMVIITGPPEAQFKAQRIFGKLKEENFPNPKEEVK
             LEAHIRVPSSTAGRVIGKGKTVNELQNLTSAEVIVPRDQTPDENEEVVRIIGHFFA
             SQTAQRKIREIVQQVKQQEEQKYPQGVASQRSK"

ORIGIN

Query Match    58.0%;    Score 1979.8;   DB 9;   Length 2010;
Best Local Similarity  99.4%;    Pred. No. 0;
Matches 1987;   Conservative  0;   Mismatches  12;   Indels  0;   Gaps  0;

Qy   35 GCCGGGGGAGCCGCGGGCTCTCGGGAAGAGACGGATGAGATGAACAAGCTTTACATCGGGA 94
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   12 GCCGTAGGGAGCCGCGGGCCTCAGGGAAGAGACGGATGAGATGAACAAGCTTTACATCGGGA 71

Qy   95 ACCTGAGCCCCGCCGTCACCGGCGACCTCCGGCAGGACCTGC 154
        |||||||||||||||||||||||||||||||||||||||||||
Db   72 ACCTGAGCCCCGCCGTCACCGGCGACCTCCGGCAGGAGCTGC 131

Qy  155 CCCTGGGGGACAGGTCCTGCTGAAGTCCGGCTACCGCGGCTACCCCGACCAGA 214
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132 CCCTGGGGGACACAGGTCCTGCTGAAGTCCGGCTACCGCGGCTACCCCGACCAGA 191

Qy  215 ACTGGGCCATCCGCGCCATCGAGACCCTCTCGGGTAAAGTGCATGGGAAAATCA 274
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  192 ACTGGGCCATCCGCGCCATCGAGACCCTCTCGGGTAAAGTGCATGGGAAAATCA 251

Qy  275 TGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGGAGCAGGAAAATTCAGATTCGAAACA 334
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  252 TGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGGAGCAGGAAAATTCAGATTCGAAACA 311

Qy  335 TCCCTCCTCACCTGCAGTGGGAGGTGTTGGATGGACTTTGGCTCAATATGGGACAGTGG 394
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  312 TCCCTCCTCACCTGCAGTGGGAGGTGTTGGATGGACTTTGGCTCAATATGGGACAGTGG 371

Qy  395 AGAATGTGGAACAAGTCAACACAGAACAGAAACCGCCGTTGTCAACGTCACATATGCAA 454
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  372 AGAATGTGGAACAAGTCAACACAGAACAGAAACCGCCGTTGTCAACGTCACATATGCAA 431

Qy  455 CAAGAGAGAAGCAAAATAGCCATGGCGGCAAGGCCACGCCGCCCGGGGCACTTCTTTAACCC 514
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  432 CAAGAGAGAAGCAAAATAGCCATGGCGGCAAGGCCACGCCGCCCGGGGCACTTCTTTAACCC 491

Qy  515 CCTTCAAGATTCCTACATCCCGGATGAAGAGGTGAGCTCCCCTTCGCCCCCTCAGCGGAG 574
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  492 CCTTCAAGATTCCTACATCCCGGATGAAGAGGTGAGCTCCCCTTCGCCCCCTCAGCGGAG 551

Qy  575 CCCAGCGTGGGGACCACTCTTCCCGGGGACCAAGGCCAAGCCACGCGCCCCGGGGGCACTTCTCAGG 634
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  552 CCCAGCGTGGGGACCACTCTTCCCGGGGACCAAGGCCAAGCCACGCGCCCCGGGGGCACTTCTCAGG 611

Qy  635 CCAGACAGATTGATTTCCCGCTGCGATCCCGGTCCCCCACCAGTTTGTTGTGCCATCA 694
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  612 CCAGACAGATTGATTTCCCGCTGCGATCCCGGTCCCCCACCAGTTTGTTGTGCCATCA 671

Qy  695 TCGAAAGGAGGGCTTGACCATAAAGAACATCACTAAGCAGACCCAGTCCGGGTAGATA 754
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db  672 TCGAAAGGAGGGCTTGACCATAAAGAACATCACTAAGCAGAACATCTAAGCAGACCCAGTCCGGGTAGATA 731

Qy  755 TCCATAGAAAGAGAGAACTCGGAGCTGGAGCTGCAGAGAGAGCCTGTCACCATCCATGCCACCCCAG 814
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  732 TCCATAGAAAGAGAGAACTCGGAGCTGGAGCTGCAGAGAGAGCCTGTCACCATCCATGCCACCCCAG 791

Qy  815 AGGGGACTTCTGAAGCATGCGCCGATGATTCTTGAAATCATGCAGAAAGAGGCAGATGAGA 874
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  792 AGGGGACTTCTGAAGCATGCGCCGATGATTCTTGAAATCATGCAGAAAGAGGCAGATGAGA 851

Qy  875 CCAAACTAGCCGAGAGATCTTGCACACAATGGCTTGGTTGGAAGAC 934
        ||||||||||||||||||||||||||||||||||||||||||||
Db  852 CCAAACTAGCCGAGAGATCTTGCAAAATCTTGGCACACAATGGCTTGGTTGGAAGAC 911

Qy  935 TGATTGGAAAAGAAGGCAGAAATTGAACATGAACATGAAACAGGGACCAAGATAA 994
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  912 TGATTGGAAAAGAAGGCAGAAATTGAACATGAAACATGAAACAGGACCAAGATAA 971

Qy  995 CAATCTCATCTTGCAGGATTTGAGCATATACAACCCGAAAGAACCATCACTGTGAAGG 1054
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  972 CAATCTCATCTTGCAGGATTTGAGCATATACAACCCGAAAGAACCATCACTGTGAAGG 1031

Qy 1055 GCACAGTTGAGGCCTGTGCCAGTGCTGAGATAGAGTGCTGCGTGGTGAGGCCT 1114
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1032 GCACAGTTGAGGCCTGTGCCAGTGCTGAGATAGAGTGCTGCGTGGTGAGGCCT 1091

Qy 1115 TTGAAAATGATATGCTGGCTGTTAACCACAAGCCAATCTGATCCCAGGGTTGAACCTCA 1174
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1092 TTGAAAATGATATGCTGGCTGTTAACCACAAGCCAATCTGATCCCAGGGTTGAACCTCA 1151

Qy 1175 GCGCACTTGGCATCTTTCACACAGGACTGTCCGTGCTATCTCCACAGGGCCCCCGCG 1234
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1152 GCGCACTTGGCATCTTTCACACAGGACTGTCCGTGCTATCTCCACAGGGCCCCCGCG 1211

Qy 1235 GAGCTCCCCCGCTGCCCCGTACCCCTTCACTACCCGGATACTTCTCCAGCC 1294
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1212 GAGCTCCCCCGCTGCCCCGTACCCCTTCACTACCCCGGATACTTCTCCAGCC 1271

Qy 1295 TGTACCCCCATCACCAGTTGGCCCGTTCCCGCTGTGGGGCCTGTGGGCGCCATCATCGGGA 1354
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1272 TGTACCCCCATCACCAGTTGGCCCGTTCCCGCTGTGGGGCCTGTGGGCGCCATCATCGGGA 1331

Qy 1355 TTGTGAATCTCTTCATCCCAGAGTTGCGCGGAGCCTTATCAAGATTGCCCCTCGCGAAGGCC 1414
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1332 TTGTGAATCTCTTCATCCCAGAGTTGCGCGGAGCCTTATCAAGATTGCCCCTCGCGAAGCAC 1391

Qy 1415 ACATCAAACAGCTGGCGAGATTGGGCGGAGATTCGCCGGGGCTCTCATCAAGATTGCCCCTGCGGAAGGCC 1474
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1392 ACATCAAACAGCTGGCGAGATTGGGCGGAGATTCGCCGGGGCTCTCATCAAGATTGCCCCTGCGGAAGGCC 1451

Qy 1475 CAGACGTCAGCGAAAGGATGGTCATCATCACCGGCGCCACCGGAAGCCCAGTTCAAGGCCC 1534
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1452 CAGACGTCAGCGAAAGGATGGTCATCATCACCGGCGCCACCGGAAGCCCAGTTCAAGGCCC 1511

Qy 1535 AGGGACCGATCTTTGGGAAACTGAAAGAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGA 1594
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1512 AGGGACCGATCTTTGGGAAACTGAAAGAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGA 1571

Qy 1595 AGCTGGAAGCGCATATCAGAGTGCCCTCTTCCACAGCTGGCCCGGGTGATTGGCAAAGGTG 1654
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1572 AGCTGGAAGCGCATATCAGAGTGCCCTCTTCCACAGCTGGCCCGGGTGATTGGCAAAGGTG 1631

Qy 1655 GCAAGACCGTGAACGAACTGCAGAACTTAACCAGTGCAGAGGTCATCGTGCCTCGTGACC 1714
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1632 GCAAGACCGTGAACGAACTGCAGAACTTAACCAGTGCAGAGGTCATCGTGCCTCGTGACC 1691

Qy 1715 AAACGCCAGATGAAATGAAGAAGTGATCGTCAGAATTATGGGACACTTCTTTGCTAGCC 1774
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1692 AAACGCCAGATGAAATGAAGAAGTGATCGTCAGAATTATGGGACACTTCTTTGCTAGCC 1751

Qy 1775 AGACTGCACAGCGCAAGATCAGAGAATTGTACAACAGGTGAAGCAGGAGCAGCAGAAAT 1834
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1752 AGACTGCACAGCGCAAGATCAGAGAATTGTACAACAGGTGAAGCAGGAGCAGCAGAAAT 1811
```

```
Qy   1835  ACCCTCAGGGAGTCGCCTCACAGCGCAGCAAGTGAGGCTCCCACAGGCACCAGCAAAACA  1894
                 ||||||||||||||||| ||||||| ||||||||||||||||||| |||||||||||||
Db   1812  ACCCTCAGGGAGTCGCCTCACAGCGCAGCAAGTGAGGCTCCCACAGGCACCAGCAAAACA  1871

Qy   1895  ACGGATGAATGTAGCCCTTCCAACACCTGACAGAATGAGACCAAACGCAGCCAGCCAGAT  1954
                 |||||||||||||||||||||||||||||||||| ||||||||| ||||||||||||||
Db   1872  ACGGATGAATGTAGCCCTTCCAACACCTGACAGAATGAGACCAAACGCAGCCAGCCAGAT  1931

Qy   1955  CGGGAGCAAACCAAAGACCATCTGAGGAATGAGAAGTCTGCGGAGGCGGCCAGGGACTCT  2014
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1932  CGGGAGCAAACCAAAGACCATCTGAGGAATGAGAAGTCTGCGGAGGCGGCCAGGGACTCT  1991

Qy   2015  GCCGAGGCCCTGAGAACCC  2033
                 |||||||||||||||||||
Db   1992  GCCGAGGCCCTGAGAACCC  2010
```

Search completed: July 21, 2004, 03:45:54
Job time : 13026 secs

```
                    GenCore version 5.1.6
              Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        July 20, 2004, 20:33:42 ; Search time 1237 Seconds
                                         (without alignments)
                                         1717.746 Million cell updates/sec

Title:              US-09-270-437D-6
Perfect score:      3412
Sequence:           1 ggcagcggaggaggcgagga........aaccttgaaatgttattt 3412

Scoring table:      IDENTITY_NUC
                    Gapop 10.0 , Gapext 1.0

Searched:           3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters:    6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database  :      N_Geneseq_29Jan04:*
              1:  geneseqn1980s:*
              2:  geneseqn1990s:*
              3:  geneseqn2000s:*
              4:  geneseqn2001as:*
              5:  geneseqn2001bs:*
              6:  geneseqn2002s:*
              7:  geneseqn2003as:*
              8:  geneseqn2003bs:*
              9:  geneseqn2003cs:*
             10:  geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	3411	100.0	3412	3	AAZ36152	Aaz36152 DNA encod
2	3320.6	97.3	3694	4	AAS26150	Aas26150 Human cDN
3	3320.6	97.3	3694	3	ABX73491	Abx73491 Human nov
4	3143	92.1	3283	3	AAZ36154	Aaz36154 An altern
5	3114.8	91.3	3667	5	AAS70982	Aas70982 DNA encod
6	3114.8	91.3	3667	6	ABS76442	Abs76442 cDNA enco
7	1999.8	58.3	2290	4	AAS26153	Aas26153 Human cDN
8	1999.8	58.3	2290	7	ABX73494	Abx73494 Human nov
9	1979.8	58.0	2010	7	AAS70981	Aas70981 DNA encod
10	1423.8	41.7	1707	7	ACA90176	Aca90176 cDNA enco
11	1096.8	32.1	1186	4	AAS26570	Aas26570 Human cDN
12	1096.8	32.1	1186	4	ABX73911	Abx73911 Human nov
13	776.2	22.7	822	4	AAS26151	Aas26151 Human cDN
14	776.2	22.7	822	7	ABX73492	Abx73492 Human nov
15	712	20.9	1740	3	AAC66035	Aac66035 Human lun
16	712	20.9	1740	6	ABL49254	Abl49254 Human lun
17	712	20.9	1740	6	ABQ92440	Abq92440 Human lun
18	712	20.9	1743	6	ABL49299	Abl49299 Human lun
19	712	20.9	1743	6	ABL49297	Abl49297 Human lun
20	712	20.9	1743	6	ABQ92483	Abq92483 Human lun
21	712	20.9	1743	6	ABQ92485	Abq92485 Human lun
22	712	20.9	1743	8	ADA28540	Ada28540 Recombina
23	712	20.9	1743	8	ADA28537	Ada28537 Recombina
24	710.4	20.8	1740	8	ADA28437	Ada28437 Human lun
25	710.4	20.8	4159	3	AAZ36150	Aaz36150 DNA encod
26	710.4	20.8	4181	6	AAC65900	Aac65900 Human lun
27	710.4	20.8	4181	6	ABL65405	Abl65405 Lung canc
28	710.4	20.8	4181	6	ABL49119	Abl49119 Human lun
29	710.4	20.8	4181	8	ABQ92305	Abq92305 Human lun
30	710.4	20.8	4181	8	ADA28265	Ada28265 Human lun
31	710.4	20.8	4181	9	ADD14661	Add14661 Human src
32	710.4	20.8	4434	9	ADE53470	Ade53470 Human lun
33	710.4	20.8	4434	8	ACH03940	Ach03940 Human cDN
34	710.4	20.8	4601	6	ABA99958	Aba99958 Human KOC
35	709	20.8	1764	6	ABL49283	Abl49283 Human lun
36	709	20.8	1764	6	ABQ92469	Abq92469 Human lun
37	707.4	20.7	1764	8	ADA28518	Ada28518 Recombina
38	690.2	20.2	2224	2	AAZ10617	Aaz10617 cDNA enco
39	643.8	18.9	1985	5	AAS76779	Aas76779 DNA encod
40	615.2	18.0	4264	3	AAS86150	Aas86150 DNA encod
41	611	17.9	710	3	AAA02565	Aaa02565 Human col
42	570.6	16.7	1946	3	AAZ36153	Aaz36153 An altern
43	489.2	14.3	1708	3	AAZ36151	Aaz36151 DNA encod
44	469.8	13.8	564	8	AAL57525	Aal57525 Human IGF
45	434	12.7	2780	4	AAK94782	Aak94782 Human ful

```
                              ALIGNMENTS

RESULT 1
AAZ36152
ID   AAZ36152 standard; DNA; 3412 BP.
XX
AC   AAZ36152;
XX
DT   11-FEB-2000  (first entry)
XX
DE   DNA encoding cancer associated antigen KOC-3.
XX
KW   Cancer associated antigen; KOC-3; cancer; vaccine; CT7; ss.
XX
OS   Homo sapiens.
XX
PN   WO995473B-A1.
XX
PD   28-OCT-1999.
XX
PF   16-MAR-1999;    99WO-US005766.
XX
PR   17-APR-1998;    98US-00061709.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Chen Y, Gure A, Tsang S, Stockert E, Jager E, Knuth A, Old LJ;
XX
DR   WPI; 2000-013284/01.
XX
PT   Nucleotides representing cancer-associated genes, used to develop
PT   products for the diagnosis, monitoring and treatment of cancers.
XX
PS   Claim 55; Page 41; 44pp; English.
XX
CC   The present sequence represents a cancer associated antigen gene
CC   designated KOC-3. The specification also describes a cancer associated
CC   antigen designated CT7. The CT7 polynucleotide was isolated from SK-MEL-
CC   37 melanoma cells. The polypeptide has some homology with MAGE-10,
CC   limited to about 210 carboxy terminal amino acids. The amino terminal of
CC   the protein has a repetitive pattern, with repeats rich in serine,
CC   proline, glutamine and leucine, and an almost invariable core of the
CC   peptide given in AAY43877. The CT7 polypeptide can be processed to
CC   peptides which provoke lysis by cytolytic T cells. The polynucleotides
CC   and polypeptides can be used for treating a cancerous condition and
CC   screening for or diagnosing cancerous conditions. The cancer associated
CC   antigens can be used as an immunogenic or vaccine composition with an
CC   adjuvant, e.g. a cytokine, a saponin, or granulocyte macrophage-colony
```

```
CC  stimulating factor (GM-CSF)
XX
SQ  Sequence 3412 BP; 970 A; 887 C; 804 G; 750 T; 0 U; 1 Other;

Query Match          100.0%;  Score 3411;  DB 3;  Length 3412;
Best Local Similarity 100.0%;  Pred. No. 0;
Matches 3412;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      1 GGCAGCCGGAGGAGGCGAGGAGGCGCCCGGTACCGGCGGCCGGGAGCCGGGGCTTCGGGG
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGCAGCCGGAGGAGGCGAGGAGGCGCCCGGTACCGGCGGCCGGGAGCCGGGGCTTCGGGG   60

Qy     61 AAGAGACGGATGATGAACAAGCTTTACATGGGGACAGAAGCTGCCCTGGGGACAGGTGTC
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAGAGACGGATGATGAACAAGCTTTACATGGGGACAGAAGCTGCCCTGGGGACAGGTGTC  120

Qy    121 GACCTTCGGACAGCTCTTTGGGGACAGCAGGAAGCTGTCCCTGGGGGACAGAGGTCCTGA
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GACCTTCGGACAGCTCTTTGGGGACAGCAGGAAGCTGTCCCTGGGGGACAGAGGTCCTGA  180

Qy    181 TCCGGCTACGGCCTTCGTGGGACTACCCCGACCAGCTGGGCCATGTGGGCCTGTCGCTGG
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCCGGCTACGGCCTTCGTGGGACTACCCCGACCAGCTGGGCCATGTGGGCCTGTCGCTGG

Qy    241 CTCTCGGGTAAAGTGGAATTGCATGGGAGAAATCATGGAGTTGATTACTCAGTCTCTAAA
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTCTCGGGTAAAGTGGAATTGCATGGGAGAAATCATGGAGTTGATTACTCAGTCTCTAAA  300

Qy    301 AAGCTAAGGAGCAGGAAAATTCAGATTCGAACACATCCCTCCTCACCTGCAGTGGGAGTG
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AAGCTAAGGAGCAGGAAAATTCAGATTCGAACACATCCCTCCTCACCTGCAGTGGGAGTG  360

Qy    361 TTGGATGGACTTTTGGCTCAATATGGACAGTGGAGAATGTGGAACAGTCAACACAGAC
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTGGATGGACTTTTGGCTCAATATGGACAGTGGAGAATGTGGAACAGTCAACACAGAC  420

Qy    421 ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAAGAAGAGCAAAAATAGCCATG
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAAGAAGAGCAAAAATAGCCATG  480

Qy    481 GAGAAGCTAAGCGGGCCATCCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGAT
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GAGAAGCTAAGCGGGCCATCCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGAT  540

Qy    541 GAAGAGGTGAGCTCCCCCTTCGCCCCCTGGGGCTGGGACCCCAGCGCCCAGCGCTCTTCCCCG
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GAAGAGGTGAGCTCCCCCTTCGCCCCCTGGGGCTGGGACCCCAGCGCCCAGCGCTCTTCCCCG  600

Qy    601 GAGCAAGGCCACGCCCCTGGGGGCCACTTCTCAGGCCAGACAGATTGATTTCCCCGTGCGG
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GAGCAAGGCCACGCCCCTGGGGGCCACTTCTCAGGCCAGACAGATTGATTTCCCCGTGCGG  660

Qy    661 ATCCTTGGTCCCCACCAGTTTGTTGGTGCCATCATCGGGCATCGGGAAAGAGGGCTTGACCATAAAG
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ATCCTTGGTCCCCACCAGTTTGTTGGTGCCATCATCGGGCATCGGGAAAGAGGGCTTGACCATAAAG  720

Qy    721 AACATCACTAAGCAGCAGCCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AACATCACTAAGCAGCAGCCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT  780

Qy    781 GCAGGAAGCCTGTCACCATCATCCCACCCACCCACCCAGGAGGGACTTCTGAAGCATGCCGCATG
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GCAGGAAGCCTGTCACCATCATCCCACCCACCCACCCAGGAGGGACTTCTGAAGCATGCCGCATG  840

Qy    841 ATTCTTGAAATCATGCAACAAATGCTTGGTTGGTGGAAAGAGTAGCCGAAGAGATTCCTCTG
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 ATTCTTGAAATCATGCAACAAATGCTTGGTTGGTGGAAAGAGTAGCCGAAGAGATTCCTCTG  900

Qy    901 AAAATCTTGCACACAAATGCTTGGTTGGATTGGAAGACTGTGTTGGAAAGAGAGGCAGAAATTTG
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 AAAATCTTGCACACAAATGCTTGGTTGGATTGGAAGACTGTGTTGGAAAGAGAGGCAGAAATTTG  960

Qy    961 AAGAAAATTGAACATGAAACATGAAACAGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AAGAAAATTGAACATGAAACATGAAACAGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020

Qy   1021 ATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGGCTGTGCCAGTGCT
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGGCTGTGCCAGTGCT 1080

Qy   1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGGAGGCTTGAAAATGATATGCTGGCTGTTAAC
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGGAGGCTTGAAAATGATATGCTGGCTGTTAAC 1140

Qy   1141 CAACAAGCCAATCTGATCCCAGGTTGAACCTCAGCGCACTTGGCATCTTTCAACAGGA
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CAACAAGCCAATCTGATCCCAGGTTGAACCTCAGCGCACTTGGCATCTTTCAACAGGA 1200

Qy   1201 CTGTCCGTGCTATCTCACCAGCAGGAGTCCCCGCGGAGTCCCCCGTGCCCCCTACCAC
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CTGTCCGTGCTATCTCACCAGCAGGAGTCCCCGCGGAGTCCCCCGTGCCCCCTACCAC 1260

Qy   1261 CCCTTCACTACCCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTGGGCCCG
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CCCTTCACTACCCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTGGGCCCG 1320

Qy   1321 TTCCCGATCATCATCCTATCCGGGACAGGAGCAGGAGATTGTGAATCTCTTCATCCCAACCCAG
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TTCCCGATCATCATCCTATCCGGGACAGGAGCAGGAGATTGTGAATCTCTTCATCCCAACCCAG 1380

Qy   1381 GCTGTGGGCGCCATCATCGGGGAAGAGGGGCACACATCAAACAGCTGGCGAGATTCGCC
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GCTGTGGGCGCCATCATCGGGGAAGAGGGGCACACATCAAACAGCTGGCGAGATTCGCC 1440

Qy   1441 GGAGCCTCATCAAGATTGCCCCTGCGAAGGCCCAGAGTCAGCGAAAGGATGGTCATC
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GGAGCCTCATCAAGATTGCCCCTGCGAAGGCCCAGAGTCAGCGAAAGGATGGTCATC 1500

Qy   1501 ATCACCGGGCCACCGGAGCCCAGTTCAAGGCCCAGCCCAGATCTTTGGGAAACTGAAA
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 ATCACCGGGCCACCGGAGCCCAGTTCAAGGCCCAGCCCAGATCTTTGGGAAACTGAAA 1560

Qy   1561 GAGGAAAACTTCTTTAACCCCAAAGAAGAAGAGAAGAAGTGAAGCTGGAGGCGCATATCGAGTGCC
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 GAGGAAAACTTCTTTAACCCCAAAGAAGAAGAGAAGAAGTGAAGCTGGAGGCGCATATCGAGTGCC 1620

Qy   1621 TCTTCCACAGCTGGCGGTGGCCCGGGTGATTGGCAAGGTCCAAGACCGTGACGAACTGCAGAAC
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TCTTCCACAGCTGGCGGTGGCCCGGGTGATTGGCAAGGTCCAAGACCGTGACGAACTGCAGAAC 1680

Qy   1681 TTAACCAGTGCAGAAGTCATCGTGACCAAGAATTCCTGACCAGCCCAGATGAAATGAGGAAGTG
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TTAACCAGTGCAGAAGTCATCGTGACCAAGAATTCCTGACCAGCCCAGATGAAATGAGGAAGTG 1740

Qy   1741 ATCGTCGAATTATCGGGCACTTCTTTGCTCAGCCTAGCGGCAGAAACAAGGACCATCTGAG
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 ATCGTCGAATTATCGGGCACTTCTTTGCTCAGCCTAGCGGCAGAAACAAGGACCATCTGAG 1800

Qy   1801 ATTGTACAACAGGTGAAGCAGCAGCAGAAATACCCTCAGGGAGTCGCCTTCACAGCGGC
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 ATTGTACAACAGGTGAAGCAGCAGCAGAAATACCCTCAGGGAGTCGCCTTCACAGCGGC 1860

Qy   1861 AGCAAGTGAGGCTCCCACAGCAGCCACCAGCAGAACAAGGATGAATGAATGTAGCCCCAACAC
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 AGCAAGTGAGGCTCCCACAGCAGCCACCAGCAGAACAAGGATGAATGAATGTAGCCCCAACAC 1920

Qy   1921 CTGACAGAATGAGACCAAACGCAGCCAGCATCGGGAGCAAACCAAGACCATCTGAG
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CTGACAGAATGAGACCAAACGCAGCCAGCATCGGGAGCAAACCAAGACCATCTGAG 1980

Qy   1981 GAATGAGAGTCTGCGAGGCGGCGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCTCAGGGGC
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GAATGAGAGTCTGCGAGGCGGCGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCTCAGGGGC 2040

Qy   2041 CGAGGAGGGCGCGGGAAGGTCAGCCAGTTTGCCGAGACCACCACCAGGTTTGCAGGATTTGAGC
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 CGAGGAGGGCGCGGGAAGGTCAGCCAGTTTGCCGAGACCACCACCAGGTTTGCAGGATTTGAGC 2100
```

```
Qy  3181 GTGTTTTTACCTCAGCACCTTGCTCTGTGTTCCCTTAGAGATTTGTAAAGCTGATAG 3240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3181 GTGTTTTTACCTCAGCACCTTGCTCTGTGTTCCCTTAGAGATTTGTAAAGCTGATAG 3240

Qy  3241 TTGGAGCATTTTTTATTTTTTAATAAAATGAGTTGGAAAAAAATAAGATATCAACT 3300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3241 TTGGAGCATTTTTTATTTTTTAATAAAATGAGTTGGAAAAAAATAAGATATCAACT 3300

Qy  3301 GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGCTTCCGCT 3360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3301 GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCT 3360

Qy  3361 AGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3412
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3361 AGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3412
```

RESULT 2
AAS26150
ID   AAS26150 standard; cDNA; 3694 BP.
XX
AC   AAS26150;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human cDNA encoding a novel secreted protein, Seq ID 329.
XX
KW   Human; immunosuppressive; antiarthritic; ss; antirheumatic; cytostatic;
KW   cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW   antibacterial; virucide; fungicide; opthalmalogical; vulnerary;
KW   secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; skin ageing; food additive;
KW   preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001341.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0186664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0209963P.
PR   26-JUL-2000; 2000US-0220664P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
```
Db  2041 CGAGGAGGGGCGGGAAGGTCAGCCAGTTGGCCAGAACCACCGAGCCGCCTCCCGCC 2100

Qy  2101 CCCCAGGGCTTCTGCAGGGTTCAGCCATCCACTTCACCATCCTCCTGAA 2160
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101 CCCCAGGGCTTCTGCAGGGTTCAGCCATCCACTTCACCATCCTCCTGAA 2160

Qy  2161 CTCCCACGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161 CTCCCACGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2220

Qy  2221 AAAATGCACACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAA 2280
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221 AAAATGCACACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAA 2280

Qy  2281 GGGAAGATGTTAAGAGATGTGGCCCTGTGGGTTACACAGGGTGCTGCAGCGGTAATATAT 2340
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281 GGGAAGATGTTAAGAGATGTGGCCCTGTGGGTTACACAGGGTGCTGCAGCGGTAATATAT 2340

Qy  2341 TTTAGAAAATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTT 2400
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341 TTTAGAAAATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTT 2400

Qy  2401 TTTTTCTTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGATAAAGTCTTTGGGAG 2460
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401 TTTTTCTTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGATAAAGTCTTTGGGAG 2460

Qy  2461 GTCTCACGGTGTAGAGAGAGCTTTGAGGCCACCCGACAAAATTCACCCAGAGGGAAAT 2520
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461 GTCTCACGGTGTAGAGAGAGCTTTGAGGCCACCCGACAAAATTCACCCAGAGGGAAAT 2520

Qy  2521 CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTCAACAGAGAGGGATA 2580
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521 CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTCAACAGAGAGGGATA 2580

Qy  2581 CCGTCTCCTGAAGAGAAACTCTGTCACTCCTCAGCTCATACACCCATT 2640
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581 CCGTCTCCTGAAGAGAAACTCTGTCACTCCTCAGCTCATACACCCATT 2640

Qy  2641 TCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCT 2700
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641 TCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCT 2700

Qy  2701 CTCTCTGTTTATCTCTCCCTCCCCTTCTTCTCCATCTCCATTCTTTTGAA 2760
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701 CTCTCTGTTTATCTCTCCCTCCCCTTCTTCTCCATCTCCATTCTTTTGAA 2760

Qy  2761 TTTCCTCATCCTCCAATCTCAATCCCGTATCTACGGCAAAAGCA 2820
         ||||||||||||||||||||||||||||||||||||||||||||
Db  2761 TTTCCTCATCCTCCAATCTCAATCCCGTATCTACGGCAAAGCA 2820

Qy  2821 GTGCTCTGAGTATCACATCACACAAAGGAACAAAAGCGAAACACACACCAGCCTCAA 2880
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821 GTGCTCTGAGTATCACATCACACAAAGGAACAAAAGCGAAACACACACCAGCCTCAA 2880

Qy  2881 CTTACACTTGGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAG 2940
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881 CTTACACTTGGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAG 2940

Qy  2941 AGGAAAACAGGAACCACCACCAAATCAACCAAACAAAGAAAAATTCCACAATG 3000
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2941 AGGAAAACAGGAACCACCACCAAATCAACCAAACAAAGAAAAATTCCACAATG 3000

Qy  3001 AAAGAATGTATTTGTCTTTTTGCATTTGGTTATAAGCTATAAAATGTCAGCAAAATG 3060
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3001 AAAGAATGTATTTGTCTTTTTGCATTTGGTTATAAGCTATAAAATGTCAGCAAAATG 3060

Qy  3061 ATTCCTTTCTTTAAAAAAATGTGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCC 3120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3061 ATTCCTTTCTTTAAAAAAATGTGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCC 3120

Qy  3121 CAGGGGCTTAAATTCACAGATTTTTTAACGGAGAAAAACACACAGAGAAGCTACCTCAG 3180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3121 CAGGGGCTTAAATTCACAGATTTTTTAACGGAGAAAAACACACAGAGAAGCTACCTCAG 3180
```

PR   14-AUG-2000;   2000US-0255757P.
PR   14-AUG-2000;   2000US-0255758P.
PR   14-AUG-2000;   2000US-0255759P.
PR   18-AUG-2000;   2000US-0262279P.
PR   22-AUG-2000;   2000US-0226681P.
PR   22-AUG-2000;   2000US-0226868P.
PR   22-AUG-2000;   2000US-0227182P.
PR   23-AUG-2000;   2000US-0227009P.
PR   30-AUG-2000;   2000US-0228924P.
PR   01-SEP-2000;   2000US-0229287P.
PR   01-SEP-2000;   2000US-0229343P.
PR   01-SEP-2000;   2000US-0229344P.
PR   01-SEP-2000;   2000US-0229345P.
PR   05-SEP-2000;   2000US-0229509P.
PR   05-SEP-2000;   2000US-0229513P.
PR   06-SEP-2000;   2000US-0230437P.
PR   06-SEP-2000;   2000US-0230438P.
PR   08-SEP-2000;   2000US-0231242P.
PR   08-SEP-2000;   2000US-0231243P.
PR   08-SEP-2000;   2000US-0231244P.
PR   08-SEP-2000;   2000US-0231413P.
PR   08-SEP-2000;   2000US-0231414P.
PR   08-SEP-2000;   2000US-0232080P.
PR   08-SEP-2000;   2000US-0232081P.
PR   12-SEP-2000;   2000US-0231968P.
PR   14-SEP-2000;   2000US-0232397P.
PR   14-SEP-2000;   2000US-0232398P.
PR   14-SEP-2000;   2000US-0232399P.
PR   14-SEP-2000;   2000US-0232400P.
PR   14-SEP-2000;   2000US-0232401P.
PR   14-SEP-2000;   2000US-0233063P.
PR   14-SEP-2000;   2000US-0233064P.
PR   14-SEP-2000;   2000US-0233065P.
PR   21-SEP-2000;   2000US-0234223P.
PR   21-SEP-2000;   2000US-0234274P.
PR   25-SEP-2000;   2000US-0234997P.
PR   25-SEP-2000;   2000US-0234998P.
PR   26-SEP-2000;   2000US-0235484P.
PR   27-SEP-2000;   2000US-0235834P.
PR   27-SEP-2000;   2000US-0235836P.
PR   29-SEP-2000;   2000US-0236327P.
PR   29-SEP-2000;   2000US-0236367P.
PR   29-SEP-2000;   2000US-0236368P.
PR   29-SEP-2000;   2000US-0236369P.
PR   29-SEP-2000;   2000US-0236370P.
PR   02-OCT-2000;   2000US-0236802P.
PR   13-OCT-2000;   2000US-0239937P.
PR   20-OCT-2000;   2000US-0240960P.
PR   20-OCT-2000;   2000US-0241221P.
PR   20-OCT-2000;   2000US-0241785P.
PR   20-OCT-2000;   2000US-0241786P.
PR   20-OCT-2000;   2000US-0241787P.
PR   20-OCT-2000;   2000US-0241808P.
PR   20-OCT-2000;   2000US-0241809P.
PR   20-OCT-2000;   2000US-0241826P.
PR   01-NOV-2000;   2000US-0244617P.
PR   08-NOV-2000;   2000US-0246474P.
PR   08-NOV-2000;   2000US-0246475P.
PR   08-NOV-2000;   2000US-0246476P.
PR   08-NOV-2000;   2000US-0246477P.
PR   08-NOV-2000;   2000US-0246478P.
PR   08-NOV-2000;   2000US-0246523P.
PR   08-NOV-2000;   2000US-0246524P.
PR   08-NOV-2000;   2000US-0246525P.
PR   08-NOV-2000;   2000US-0246526P.
PR   08-NOV-2000;   2000US-0246527P.
PR   08-NOV-2000;   2000US-0246528P.
PR   08-NOV-2000;   2000US-0246532P.

PR   08-NOV-2000;   2000US-0246609P.
PR   08-NOV-2000;   2000US-0246610P.
PR   08-NOV-2000;   2000US-0246611P.
PR   08-NOV-2000;   2000US-0246613P.
PR   17-NOV-2000;   2000US-0249207P.
PR   17-NOV-2000;   2000US-0249208P.
PR   17-NOV-2000;   2000US-0249209P.
PR   17-NOV-2000;   2000US-0249210P.
PR   17-NOV-2000;   2000US-0249211P.
PR   17-NOV-2000;   2000US-0249212P.
PR   17-NOV-2000;   2000US-0249213P.
PR   17-NOV-2000;   2000US-0249214P.
PR   17-NOV-2000;   2000US-0249215P.
PR   17-NOV-2000;   2000US-0249216P.
PR   17-NOV-2000;   2000US-0249217P.
PR   17-NOV-2000;   2000US-0249218P.
PR   17-NOV-2000;   2000US-0249244P.
PR   17-NOV-2000;   2000US-0249245P.
PR   17-NOV-2000;   2000US-0249264P.
PR   17-NOV-2000;   2000US-0249265P.
PR   17-NOV-2000;   2000US-0249297P.
PR   17-NOV-2000;   2000US-0249299P.
PR   17-NOV-2000;   2000US-0249300P.
PR   01-DEC-2000;   2000US-0250016P.
PR   01-DEC-2000;   2000US-0250391P.
PR   05-DEC-2000;   2000US-0251030P.
PR   05-DEC-2000;   2000US-0251198P.
PR   05-DEC-2000;   2000US-0256719P.
PR   06-DEC-2000;   2000US-0251479P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
PR   08-DEC-2000;   2000US-0251989P.
PR   08-DEC-2000;   2000US-0251990P.
PR   11-DEC-2000;   2000US-0254097P.
PR   05-JAN-2001;   2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,   Barash SC,   Ruben SM;
XX
DR   WPI; 2001-488783/53.
DR   P-PSDB; AAU16163.
XX
PT   New nucleic acid molecules encoding 461 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives.
XX
PS   Claim 1; SEQ ID NO 329; 980pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Antibodies to the proteins can also be used in
CC   alleviating symptoms associated with the disorders and in diagnostic
CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC   and many other disorders listed in the specification. The polypeptides
CC   can also be used to aid wound healing and epithelial cell proliferation,
CC   to prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities; fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence encodes a novel secreted protein of the invention. Note: The

```
CC  sequence data for this patent did not form part of the printed

Query Match       97.3%;   Score 3320.6;   DB 4;   Length 3694;
Best Local Similarity   99.9%;   Pred. No. 0;
Matches 3343;  Conservative 1;  Mismatches 1;  Indels 3;  Gaps 2;

Qy    68 GGATGATGAACAAGCTTACATCGGGAACTGAGCCCCGCCGTCACCGCGACGACCTCC 127
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    91 GGATGATGAACAAGCTTTACATCGGGAACTGAGCCCCGCCGTCACCGCGACGACCTCC 150

Qy   128 GGCAGCTCTTGGGACAGGAGTCGCCCTGCCCTGGGGGACAGGTCCTGCGAAGTCCGGCT 187
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   151 GGCAGCTCTTGGGACAGGAGTCGCCCTGCCCTGGGGGACAGGTCCTGCGAAGTCCGGCT 210

Qy   188 ACGGCCTTCGTGGACTACCCCGACCAGAACTGGGCCATCGCGCCATCGCGCGCCTCTCGG 247
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   211 ACGGCCTTCGTGGACTACCCCGATACTTCTCCAGCGTGCCCATCGCGCGCCTCTCGG 270

Qy   248 GTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTTCTAAAAAGCTAA 307
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   271 GTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTTCTAAAAAGCTAA 330

Qy   308 GGAGCAGGAAAATTCAGATTCGAAACATCCCTCCACCTGCAGTGGGAGTGTTGGATG 367
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   331 GGAGCAGGAAAATTCAGATTCGAAACATCCCTCCACCTGCAGTGGGAGTGTTGGATG 390

Qy   368 GACTTTTGGCTCAATATGGACAGTTGGAAATGTGAACAAGTCAACACAGACACAGAAA 427
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   391 GACTTTTGGCTCAATATGGACAGTTGGAAATGTGAACAAGTCAACACAGACACAGAAA 450

Qy   428 CCGCCGTTGTCAACGTCACATATGCAACACATATGCAACAAGAAGAAGCAAAAATAGCC 487
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   451 CCGCCGTTGTCAACGTCACATATGCAACACATATGCAACAAGAAGAAGCAAAAATAGCC 510

Qy   488 TAAGCGGGCATCAGTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGATGAAGAGG 547
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   511 TAAGCGGGCATCAGTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGATGAAGAGG 570

Qy   548 TGAGTCCCCTTCGCCCCCTCAGCGAGCCAGCCGTGGGGACCACTCTTCCCGGAGCCAAG 607
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   571 TGAGTCCCCTTCGCCCCCTCAGCGAGCCAGCCGTGGGGACCACTCTTCCCGGAGCCAAG 630

Qy   608 GCCAGCGCCCCGGGGCCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGGATCCTGG 667
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   631 GCCAGCGCCCCGGGGCCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGGATCCTGG 690

Qy   668 TCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGAGGAGCCTTGACCATAAAGAACATCA 727
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   691 TCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGAGGAGCCTTGACCATAAAGAACATCA 750

Qy   728 CTAAGCAGAACCAGTCCCGGTAGATATCCAGAAAGAGAACTCTGGAGCTGCAGAGA 787
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   751 CTAAGCAGAACCAGTCCCGGTAGATATCCAGAAAGAGAACTCTGGAGCTGCAGAGA 810

Qy   788 AGCCTGTCACCATCCATCCATCCCAGAGGGGACTTCTGAAGCATGCCGCCATGATTCTTG 847
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   811 AGCCTGTCACCATCCATCCATCCCAGAGGGGACTTCTGAAGCATGCCGCCATGATTCTTG 870

Qy   848 AAATCATGCAGAAAGAGGCAGATGACCAAACTAGCCGAAGAGATTCCTCTGAAAATCT 907
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   871 AAATCATGCAGAAAGAGGCAGATGACCAAACTAGCCGAAGAGATTCCTCTGAAAATCT 930

Qy   908 TGGCACACAAATGGCTTGGTTGGAAGACTGATTGGAAGAGCAGAAATTGAAGAAAA 967
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   931 TGGCACACAAATGGCTTGGTTGGAAGACTGATTGGAAGAGCAGAAATTGAAGAAAA 990

Qy   968 TTGAACATGAAACAGGAGACCAAGATAACAATTCCATCTTTGCAGGATTTGAGCATATACA 1027
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   991 TTGAACATGAAACAGGAGACCAAGATAACAATTCCATCTTTGCAGGATTTGAGCATATACA 1050

Qy  1028 ACCCGGAAAGAACCATCACTGTGAGGGCACCAGTTGAGGCTGCCAGTGCCAGTGAGATAG 1087
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1051 ACCCGGAAAGAACCATCACTGTGAGGGCACCAGTTGAGGCTGCCAGTGCCAGTGAGATAG 1110

Qy  1088 AGATTATGAAGAAGCTGCGTGAGGCCTTGAAAATGATATGCTGGCTGTTAACCAACAAG 1147
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1111 AGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTAACCAACAAG 1170

Qy  1148 CCAATCTGATCCCGACCTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGACTGTCCG 1207
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1171 CCAATCTGATCCCGACCTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGACTGTCCG 1230

Qy  1208 TGCTATCTCCACCAGCAGGAGCCTCCCCGGCCCCCGCCTGCCCCCGCCTGCCCCCCTTCA 1267
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1231 TGCTATCTCCACCAGCAGGAGCCTCCCCGGCCCCCGCCTGCCCCCGCCTGCCCCCCTTCA 1290

Qy  1268 CTACCCACTCCGGATACTTCTCCAGCAGGAGATTGTACCCGTGTACCCCCATCACCACCG 1327
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1291 CTACCCACTCCGGATACTTCTCCAGCAGGAGATTGTACCCGTGTACCCCCATCACCACCG 1350

Qy  1328 ATCATCACTCTTATCCAGCAGGAGAGATTGTGAATCTCTTCATCCCAACCCAGGCTGTGG 1387
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1351 ATCATCACTCTTATCCAGCAGGAGAGATTGTGAATCTCTTCATCCCAACCCAGGCTGTGG 1410

Qy  1388 GCGCCATCATCGGGAAGAAGGGGGCACACATCAAACAGCTGGCGGAGATTCGCCGGAGCCT 1447
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1411 GCGCCATCATCGGGAAGAAGGGGGCACACATCAAACAGCTGGCGGAGATTCGCCGGAGCCT 1470

Qy  1448 CTATCAAGATTGCCCCTGCGGAGGCCCAGCGTCAGCGCGAAGCTCAGCGAAGGATG 1507
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1471 CTATCAAGATTGCCCCTGCGGAGGCCCAGCGTCAGCGCGAAGGATGGTCATCATCACCG 1530

Qy  1508 GGCCACCGAAGCCCAGTTCAAGGCCCAGGCCCCAGGGCCGGATCTTTGGGAAACTGAAGAGAAA 1567
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1531 GGCCACCGAAGCCCAGTTCAAGGCCCAGGCCCCAGGGCCGGATCTTTGGGAAACTGAAGAGAAA 1590

Qy  1568 ACTTCTTTAACCCCAAAGAGAGTGAAGCGCATATCAGAGTGCCCCTTCCA 1627
         ||||||||||||||||||||||||||||||||||: ||||||||||||||
Db  1591 ACTTCTTTAACCCCAAAGAGAGTGAASCGCATATCAGAGTGCCCCTTCCA 1650

Qy  1628 CAGCTGGCGCCGGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTGAACTTAACCA 1687
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1651 CAGCTGGCGCCGGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTGAACTTAACCA 1710

Qy  1688 GTGCAGAGTCATCGTGCTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTGATCGTCA 1747
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1711 GTGCAGAGTCATCGTGCTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTGATCGTCA 1770

Qy  1748 GAATTATCGGGCACTTCTTGTTGTCTGCTAGCAGCAGAAATACCCTCAGGGAGTACAGCGCAGCAACT GTAC 1807
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1771 GAATTATCGGGCACTTCTTGTTGTCTGCTAGCAGCAGAAATACCCTCAGGGAGTCGCCAGCAAGT 1830

Qy  1808 AACAGGTGAAGCAGCAGGAGCAGGAGCAGAAATACCCTGAGAACCCCAGCGGGGCCGAGGAG 1867
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1831 AACAGGTGAAGCAGCAGGAGCAGGAGCAGAAATACCCTGAGAACCCCAGCGGGGCCGAGGAG 1890

Qy  1868 GAGGCTCCCACAGGCGAAGCATGCGGCAAAACACGGATGAATGTAGCCCTTCCAACACCTGACAG 1927
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1891 GAGGCTCCCACAGGCGAAGCATGCGGCAAAACACGGATGAATGTAGCCCTTCCAACACCTGACAG 1950

Qy  1928 AATGAGACCAAACGCAGCGAGCAAACCAGTCCTGAGAATAG 1987
         |||||||||||||||||||||||||||||||||||||||||
Db  1951 AATGAGACCAAACGCAGCGAGCAAACCAGTCCATCTGAGGAATGAG 2010

Qy  1988 AAGTCTGCGCGGGGAAGGTCAGCCAGTTTTGCCAGAACCCCAGGGGCCGGAGGAG 2047
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2011 AAGTCTGCGCGGGGAAGGTCAGCCAGTTTTGCCAGAACCCCAGGGGCCGGAGGAG 2070

Qy  2048 GGGCGGGAAGGTCAGCTTCAGCGAGATCTCTCGGATCTCTCCCGCCTCCCCAGG 2107
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2071 GGGCGGGAAGGTCAGCTTCAGCGAGATCTCTCGGATCTCTCCCGCCTCCCCAGG 2130

Qy  2108 GCTTCTGCGAGGCTTCAGCAGCTCTGCCAGTCGTTGCAGCCATCCATCACTCTCCTGAACTCCCAC 2167
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2131 GCTTCTGCGAGGCTTCAGCAGCTCTGCCAGTCGTTGCAGCCATCCATCACTCTCCTGAACTCCCAC 2190
```

```
Qy  2168  GACGCTATCCCTTTAGTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGC  2227
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2191  GACGCTATCCCTTTAGTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGC  2250

Qy  2228  ACACCCTTTTCTGTGGCAAATGTCTCTGTACATGTGTGTACATATTAGAAAGGGAAGA  2287
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2251  ACACCCTTTTCTGTGGCAAATGTCTCTGTACATGTGTGTACATATTAGAAAGGGAAGA  2310

Qy  2288  TGTTAAGAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCCGTAATATATTTAGAA  2347
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2311  TGTTAAGAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCCGTAATATATTTAGAA  2370

Qy  2348  ATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTATTTTTTTTTCT  2407
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2371  ATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTTTTTTTCT  2430

Qy  2408  TTTTAAAGAGAAAGCAGGCTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTCAC  2467
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2431  TTTTAAAGAGAAAGCAGGCTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTCAC  2490

Qy  2468  GGTGTAGAGAGGAGCTTTGAGGCCACCCCGACAAAATTCACCCAGAGGGAAATCTCGTCG  2527
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2491  GGTGTAGAGAGGAGCTTTGAGGCCACCCCGACAAAATTCACCCAGAGGGAAATCTCGTCG  2550

Qy  2528  GAAGGACACTCACGGCAGTTCTCGGATCACCTGTGATGTCAACAGAGGGATACCGTCTC  2587
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2551  GAAGGACACTCACGGCAGTTCTCGGATCACCTGTGATGTCAACAGAGGGATACCGTCTC  2610

Qy  2588  CTTGAAGAGAAACTCTGTCACTCCTCATGCCTGTCTAGCTCATACACCCATTTCTCTTT  2647
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2611  CTTGAAGAGAAACTCTGTCACTCCTCATGCCTGTCTAGCTCATACACCCATTTCTCTTT  2670

Qy  2648  GCTTCACAGGTTTTAAACTGGTTTTTTTGCATACTGCTATATAATCTCTGTCTCTCTG  2707
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2671  GCTTCACAGGTTTTAAACTGGTTTTTTTGCATACTGCTATATAATTCTCTGTCTCTCTG  2730

Qy  2708  TTTATCTCTCCCCCTCCCCCTCCCCTCCCCTTCTTCCATCTCCATTCTTTTGAATTCCTC  2767
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2731  TTTATCTCTCCCCCTCCCCCTCCCCTCCCCTTCTTCCATCTCCATTCTTTTGAATTCCTC  2790

Qy  2768  ATCCCTCCATCTCAATCCGTATCCACCA--CCCCCCCCCCCCCCCAGGCAAAGCAGTTGCT  2825
          |||||||||||||||||||||||||||||          ||||||||||||||||||||
Db  2791  ATCCCTCCATCTCAATCCGTATCCACCACCCCCCCCCCCCCCCCCAGGCAAAGCAGTTGCT  2850

Qy  2826  CTGAGTATCACATCACACAAAGGAACAAAGCCAAACACACAACCAGCCTCAACTTAC  2885
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2851  CTGAGTATCACATCACACAAAGGAACAAAGCCAAACACACAACCAGCCTCAACTTAC  2910

Qy  2886  ACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGGAA  2945
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2911  ACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGGAA  2970

Qy  2946  AACAGGAACCCACCAACCAACCAATCACCAACAAAGAAAAATTCCACAATGAAAGA  3005
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2971  AACAGGAACCCACCAACCAACCAATCACCAACAAAGAAAAATTCCACAATGAAAGA  3030

Qy  3006  ATGTATTTTGTCTTTTTGCATTTGGTATAAGCCATCAATATTCAGCAAAATGATTCC  3065
          |||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db  3031  ATGTATTTTGTCTTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCC  3090

Qy  3066  TTTCTTT-AAAAAAAAAAAAAATGTGGAGGAAGTAGAAATTTACCAAGGTGTTGGCCCAGG  3124
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3091  TTTCTTTAAAAAAAAAAAAAAAATGTGGAGGAAGTAGAAATTTACCAAGGTGTTGGCCCAGG  3150

Qy  3125  GCGTTAAATTCACAGATTTTTTTAACGAGAAAAACACAGAGAAGCTACCTCAGTGT  3184
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3151  GCGTTAAATTCACAGATTTTTTTTAACGAGAAAACACACAGAGAAGCTACCTCAGTGT  3210

Qy  3185  TTTTACCTTCAGCCACCCTTGCTCTGTGTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGG  3244
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3211  TTTTACCTTCAGCCACCCTTGCTCTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGG  3270

Qy  3245  AGCATTTTTTATTTTTTTAATAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCA  3304
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3271  AGCATTTTTTATTTTTTTAATAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCA  3330

Qy  3305  GCCTGGAGAAGGTGACAGTCCAAGTGTCCAACAGCTGTTCGAATTGTCTTCCGCTAGCC  3364
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3331  GCCTGGAGAAGGTGACAGTCCAAGTGTCCAACAGCTGTTCGAATTGTCTTCCGCTAGCC  3390

Qy  3365  AAGAACCNATATGGCCTTCTTTTGGACAACCTTGAAAATGTTATT  3412
          ||||||| ||||||||||||||||||||| ||||||||||||||
Db  3391  AAGAACCTATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTATT  3438
```

RESULT 3
ABX73491
ID   ABX73491 standard; DNA; 3694 BP.
XX
AC   ABX73491;
XX
DT   18-MAR-2003 (first entry)
XX
DE   Human novel polynucleotide #319.
XX
KW   Human; gene; ds; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW   haemostatic; antiarteriosclerotic.
XX
OS   Homo sapiens.
XX
PN   US2002132753-A1.
XX
PD   19-SEP-2002.
XX
PF   17-JAN-2001; 2001US-00764864.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180528P.
PR   28-JUN-2000; 2000US-0214886P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   22-AUG-2000; 2000US-0226868P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   08-SEP-2000; 2000US-0231413P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   27-SEP-2000; 2000US-0235834P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.

PR   29-SEP-2000;   2000US-0236370P.
PR   02-OCT-2000;   2000US-0236802P.
PR   02-OCT-2000;   2000US-0237037P.
PR   02-OCT-2000;   2000US-0237038P.
PR   02-OCT-2000;   2000US-0237039P.
PR   02-OCT-2000;   2000US-0237040P.
PR   13-OCT-2000;   2000US-0239935P.
PR   20-OCT-2000;   2000US-0240940P.
PR   20-OCT-2000;   2000US-0241785P.
PR   20-OCT-2000;   2000US-0241809P.
PR   01-NOV-2000;   2000US-0244617P.
PR   17-NOV-2000;   2000US-0249299P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA, Ruben SM, Barash SC;
XX
DR   WPI; 2003-147444/14.
DR   P-PSDB; ABU55231.
XX
PT   New polypeptides and nucleic acids, useful in gene therapy for treating,
PT   inhibiting or preventing neural e.g. neural, immune system, muscular,
PT   respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT   renal disorders.
XX
PS   Claim 1; SEQ ID NO 329; 402pp; English.
XX
CC   The invention relates to human novel polypeptides and their associated
CC   polynucleotides. The polypeptides and polynucleotides are useful in gene
CC   therapy for treating, inhibiting or preventing neural disorders, immune
CC   system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC   and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC   nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC   gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC   (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC   heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC   renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC   leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC   appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC   related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC   infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent
CC   human novel polynucleotides of the invention
XX
SQ   Sequence 3694 BP; 1103 A; 934 C; 839 G; 817 T; 0 U; 1 Other;

Query Match   97.3%;   Score 3320.6;   DB 7;   Length 3694;
Best Local Similarity   99.9%;   Pred. No. 0;
Matches 3343;   Conservative   1;   Mismatches   1;   Indels   3;   Gaps   2;

Qy   68 GGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCCGCGTCACGGCCGACGACCTCC 127
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   91 GGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCCGCGTCACGGCCGACGACCTCC 150

Qy   128 GGCAGGCTCTTTGGGACAGGAGCTTGCCCCTGTGCGCGGGACAGGTCCTGCTGAAGTCCGGCT 187
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   151 GGCAGGCTCTTTGGGACAGGAGCTTGCCCCTGTGCGCGGGACAGGTCCTGCTGAAGTCCGGCT 210

Qy   188 ACGCCTTCGTCGACTACCCCGGGAAACATCCTCCCGCGCCGCCATCGGGACCTCCCCGCTCG 247
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   211 ACGCCTTCGTCGACTACCCCGGGAAACATCCTCCCGCGCCGCCATCGGGACCCCTCCTCGG 270

Qy   248 GTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTAA 307
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   271 GTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTAA 330

Qy   308 GGAGCAGGAAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGAGTGTTGGATG 367
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   331 GGAGCAGGAAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGAGTGTTGGATG 390

Qy   368 GACTTTTGGCTCAATATGGACAGTGGAGAATGTGGA...AGTCAACACAGCAGACACAGAAA 427
            ||||||||||||||||||||||||||||| || ||| |||||||||||||
Db   391 GACTTTTGGCTCAATATGGACAGTGGAGAATGTGGA...AATAGCCATGGAGAGAGC 450

Qy   428 CCGCCGTTGTCAACGTCACATATGCAACAGAGAGAGAGC...AGTCAACACAGCAGACACAGAAA 487
            ||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   451 CCGCCGTTGTCACGTCACATATGCAACAGAGAGAGAG...AATAGCCATGGAGAGAGC 510

Qy   488 TAAGCGGGCATCAGTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGATGAAGAGG 547
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   511 TAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGATGAAGAGC 570

Qy   548 TGAGCTCCCCTTGCCCCCTCGCCCCCTCAGCGAGCCCAGGCTGGGGACCACTCTTCCCGG 607
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   571 TGAGCTCCCCTTGCCCCCTCGCCCCCTCAGCGAGCCCAGGCTGGGGACCACTCTTCCCGG 630

Qy   608 GCCACGCCCCTGGGGCACTTCTCAGCGCAGACAGATTGATTTCCCGCTGCGGATCCTGG 667
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   631 GCCACGCCCCTGGGGCACTTCTCAGCGCAGACAGATTGATTTCCCGCTGCGGGATCCTCG 690

Qy   668 TCCCCACCCAGTTGTTGGTGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAACATCA 727
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   691 TCCCCACCCAGTTGTTGGTGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAACATCA 750

Qy   728 CTAAGCAGACCCAGTCCCGGTAGATATCCATAGAAAAGAACTCTGGAGCTGCAGAGA 787
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   751 CTAAGCAGACCCAGTCCCGGTAGATATCCATAGAAAAGAACTCTGGAGCTGCAGAGA 810

Qy   788 AGCCTGTCACCATCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATGATTCTTG 847
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   811 AGCCTGTCACCATCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATGATTCTTG 870

Qy   848 AAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAATCT 907
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   871 AAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAAATCT 930

Qy   908 TGGCACAAATGCTTGGTTGGAAGACTGATTGGAAAGAAGGCCAGAAATTTGAAGAAAA 967
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   931 TGGCACAAATGGCTTGGTTGGAAGACTGATTGGAAGACTGATTGGAAAGGCAGAAATTTGAAGAAAA 990

Qy   968 TTGAACATGAAACAGGGACCAAGATAACAATCCATCTTTGCAGGATTTGAGCATATACA 1027
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   991 TTGAACATGAAACAGGGACCAAGATAACAATCCATCTTTGCAGGATTTGAGCATATACA 1050

Qy   1028 ACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGAGATAG 1087
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1051 ACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGAGATAG 1110

Qy   1088 AGATTATGAAGAAGCTGCGTGAGGGCTTTGAAAATGATATGCTGGCTGTTAACCAACAAG 1147
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1111 AGATTATGAAGAAGCTGCGTGAGGGCTTTGAAAATGATATGCTGGCTGTTAACCAACAAG 1170

Qy   1148 CCAATCTGATCCCAGGGTTGAACCTCAGGGCACTTGGCATCTTTTCAACAGGACTGTCCG 1207
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1171 CCAATCTGATCCCAGGGTTGAACCTCAGGGCACTTGGCATCTTTTCAACAGGACTGTCCG 1230

Qy   1208 TGCTATCTCCACCAGCAGGGACCTGCCCCGCGGAGCTCCCCCGCTGCCCCGTACCACCCTTCA 1267
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1231 TGCTATCTCCACCAGCAGGGACCTGCCCCGCGGAGCTCCCCCGCTGCCCCGTACCACCCCTTCA 1290

Qy   1268 CTACCCACTCCGGATACTTCTCCAGCTGTACCCCATCACCAGTTGCCCGTTCCCGC 1327
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1291 CTACCCACTCCGGATACTTCTCCAGCTGTACCCCATCACCAGTTGCCCGTTCCCGC 1350

Qy   1328 ATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTCTTCAGTCTCTAAAAGCTAA 1387
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1351 ATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTCTTCAGTCTCTTAAAAGCTAA 1410

Qy   1388 GGCGCCATCATCGGGAAGAAGAGGGGGCACACATCAAACAGCTGCCGGAGCCT 1447
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1411 GCGCCATCATCGGGGAAGAATTGCGCCGAGATTGCGGGAGATTGCGCCGAGCCT 1470

```
Qy   1448 CTATCAAGAATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCATCACCG 1507
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1471 CTATCAAGAATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCATCACCG 1530

Qy   1508 GGCCACCGGCGAAGCCCAGTTCAAGGGCCCAGGGACGGATCTTTGGGAAACTGAAGAGGAAA 1567
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1531 GGCCACCGGCGAAGCCCAGTTCAAGGGCCCAGGGACGGATCTTTGGGAAACTGAAGAGGAAA 1590

Qy   1568 ACTTCTTTAACCCCAAAGAAGAAGTGAAGCGCATATCAGAGTGCCCTCTTCCA 1627
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1591 ACTTCTTTAACCCCAAAGAAGAAGTGAAGCGGAAGCGCATATCAGAGTGCCCTCTTCCA 1650

Qy   1628 CAGCTGCGCCGGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTGCAGAACTTAACCA 1687
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1651 CAGCTGCGCCGGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTGCAGAACTTAACCA 1710

Qy   1688 GTGCAGAAGTCATCGTGCCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTGATCGTCA 1747
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1711 GTGCAGAAGTCATCGTGCCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTGATCGTCA 1770

Qy   1748 GAATTATCGGGCACTTCTTTGCTCAGCGCAGCTGCACAGCGCAAGATCAGGGAAATTGTAC 1807
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1771 GAATTATCGGGCACTTCTTTGCTCAGCGCAGCTGCACAGCGCAAGATCAGGGAAATTGTAC 1830

Qy   1808 AACAGGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAGCAAGT 1867
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1831 AACAGGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAGCAAGT 1890

Qy   1868 GAGGCTCCCACAGCACCAGCCGATGAATGTAGCCCTTCCAACACCTGACAG 1927
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1891 GAGGCTCCCACAGCACCAGCCGATGAATGTAGCCCTTCCAACACCTGACAG 1950

Qy   1928 AATGAGACCAAAACGAGCGAGCAGGGATCTCGCCAGATCCAAAGACCATCTGAGAATGAG 1987
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1951 AATGAGACCAAAACGAGCGAGCAGGGATCTCGCCAGATCCAAAGACCATCTGAGAATGAG 2010

Qy   1988 AAGTCTGCGGAGGCGGCGCAGGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGCCGAGGAG 2047
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2011 AAGTCTGCGGAGGCGGCGCAGGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGCCGAGGAG 2070

Qy   2048 GGGCGGGGAAGGTCAGCCAGGTTTGCCAGAACCACCCCGGAGCCCCGCCCTCCCGCCCCAGG 2107
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2071 GGGCGGGGAAGGTCAGCCAGGTTTGCCAGAACCACCCCGGAGCCCCGCCCTCCCGCCCCAGG 2130

Qy   2108 GCTTCTGCAGGCTTCAGCCATCCACTTCACCATCGGATCTCCTCCTGAACTCCCAC 2167
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2131 GCTTCTGCAGGCTTCAGCCATCCACTTCACCATCGGATCTCCTCCTGAACTCCCAC 2190

Qy   2168 GACGGCTATCCCTTTTAGTTAGTTGAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGC 2227
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2191 GACGGCTATCCCTTTTAGTTAGTTGAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGC 2250

Qy   2228 ACACCCTTTTTCTGTGTGGGGTTACACAGGGTGCCTGCACATGTGTACATATTATTAGAA 2287
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2251 ACACCCTTTTTCTGTGTGGGGTTACACAGGGTGCCTGCACATGTGTACATATTATTAGAA 2310

Qy   2288 TGTTAAGATATGTGCCTGTGTGGGTTACACAGGGTGCCTGCACATGTGTAATAATATTAAT 2347
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2311 TGTTAAGATATGTGCCTGTGTGGGTTACACAGGGTGCCTGCACATGTGTAATAATATTAAT 2370

Qy   2348 ATAATATATCAAATAACTAACTCCAATTTTAATCAATTATTAATTTTTTTTTTTTCT 2407
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2371 ATAATATATCAAATAACTAACTCCAATTTTAATCAATTATTAATTTTTTTTTTTTCT 2430

Qy   2408 TTTTAAAGAGAAAGCAGCAGCTTTTCTTAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTCAC 2467
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2431 TTTTAAAGAGAAAGCAGCAGCTTTTCTTAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTCAC 2490

Qy   2468 GGTGTAGAGAGGAGCTTTGAGGCCACCCGCACCAAATTCACCCAGAGGGAAATCTCGTCG 2527
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2491 GGTGTAGAGAGGAGCTTTGAGGCCACCCGCACCAAATTCACCCAGAGGGAAATCTCGTCG 2550

Qy   2528 GAAGGACACTCACCGCAGTTCTGGATCACCGTGTATGTCAACAGAGAGAGGATACCGTCTC

Db   2551 GAAGGACACTCACCGCAGTTCTGGATCACCGTGTATGTCAACAGAGAGGATACCCATTCTCTTT 2610

Qy   2588 CTTGAAGAGGAAACTCTGCCACTCCTGCCTCAGCTGCATCCTCAGCTCATACACCCATTTCTCTTT 2647
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2611 CTTGAAGAGGAAACTCTGTCACTCCTGTCTCAGCTGCATCCTCAGCTCATACACCCATTTCTCTTT 2670

Qy   2648 GCTTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCTCCTG 2707
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2671 GCTTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCTCCTG 2730

Qy   2708 TTTATCTCTCCCCTCCCCTTCTCTCCATCTCCATCTCTTTGAATTCCTC 2767
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2731 TTTATCTCTCCCCTCCCCTTCTCTCCATCTCCATCTCTTTGAATTCCTC 2790

Qy   2768 ATCCCTCCATCTCAATCCCGTATCTAGCA--CCCCCCCCCCCCCCCCAGGCAAAGCAGTGCT 2825
          |||||||||||||||||||||||||||||  |||||||||||||||||||||||||||||
Db   2791 ATCCCTCCATCTCAATCCCGTATCTAGCACCACCCCCCCCCCCCCCCAGGCAAAGCAGTGCT 2850

Qy   2826 CTGAGTATCACATCACACAAAAGGAACAAAAGCGAAACACACAAACCAGCCTCCAACTTAC 2885
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2851 CTGAGTATCACATCACACAAAAGGAACAAAAGCGAAACACACAAACCAGCCTCCAACTTAC 2910

Qy   2886 ACTTGGTTACTCAAAGAACAAGAGTCAATGGTGTCCTAGCGTTTTGGAAGAGGAA 2945
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2911 ACTTGGTTACTCAAAGAACAAGAGTCAATGGTGTACTTGTCCTAGCGTTTTGGAAGAGGAA 2970

Qy   2946 AACAGGAACCCCACCAACCAACCAATCAACCAAAGAAAAATTCCACAATGAAAGA 3005
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2971 AACAGGAACCCCACCAACCAACCAATCAACCAATCAACAGAAAAATTCCACAATGAAAGA 3030

Qy   3006 ATGTATTTTGTCTTTTGTCATTTTGGTGTGTATAAGCCATCAATAATTCAGCAAAATGATTCC 3065
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3031 ATGTATTTTGTCTTTTGTCATTTTGGTGTGTATAAGCCATCAATATTCAGCAAAATGATTCC 3090

Qy   3066 TTTCTTT-AAAAAAAAATGTGGAGAAAGTAGAATTTACCAAGGTTGTTGGCCCAGG 3124
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   3091 TTTCTTTTAAAAAAAAATGTGGAGAAAGTAGAATTTACCAAGGTTGTTGGCCCAGG 3150

Qy   3125 GGTTAAATTCACAGATTTTTTAACGAGAAAACACACAGAGAAGCTACCTCAGGTGT 3184
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3151 GGTTAAATTCACAGATTTTTTAACGAGAAAACACACAGAGAAGCTACCTCAGTGT 3210

Qy   3185 TTTTACCTCAGCACCTTGCTCTTGCCTCGTGTTCCCTTAGAGATTTGTAAAGCTGATAGTTGG 3244
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3211 TTTTACCTCAGCACCTTGCTCTTGCCTCGTGTTTCCCTTAGAGATTTGTAAAGCTGATAGTTGG 3270

Qy   3245 AGCATTTTTTTATTTTTAATAAAAAATGAGTTGAAAAATAAGATATCAACTGCCA 3304
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3271 AGCATTTTTTTATTTTTAATAAAAAATGAGTTGGAAAAATAAGATATCAACTGCCA 3330

Qy   3305 GCCTGGAGAAGGTGACAGTCCAAGTGCAACAGCTGTTCTGAATTGTCTTCCGCTAGCC 3364
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3331 GCCTGGAGAAGGTGACAGTCCAAGTGCAACAGCTGTTCTGAATTGTCTTCCGCTAGCC 3390

Qy   3365 AAGAACCNATATGGCCCTTCTTTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3412
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3391 AAGAACCTATATGGCCCTTCTTTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3438
```

```
XX
PN   WO9954738-A1.
XX
PD   28-OCT-1999.
XX
PF   16-MAR-1999;   99WO-US005766.
XX
PR   17-APR-1998;   98US-00061709.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Chen Y, Gure A, Tsang S, Stockert E, Jager E, Knuth A, Old LJ;
XX
DR   WPI; 2000-013284/01.
XX
PT   Nucleotides representing cancer-associated genes, used to develop
PT   products for the diagnosis, monitoring and treatment of cancers.
XX
PS   Claim 55; Page 42-43; 44pp; English.
XX
CC   The present sequence represents an alternative form of a cancer
CC   associated antigen gene designated KOC-3. The specification also
CC   describes a cancer associated antigen designated CT7. The CT7
CC   polynucleotide was isolated from SK-MEL-37 melanoma cells. The
CC   polypeptide has some homology with MAGE-10, limited to about 210 carboxy
CC   terminal amino acids. The amino terminal of the protein has a repetitive
CC   pattern, with repeats rich in serine, proline, glutamine and leucine, and
CC   an almost invariable core of the peptide given in AAY43877. The CT7
CC   polypeptide can be processed to peptides which provoke lysis by cytolytic
CC   T cells. The polynucleotides and polypeptides can be used for treating a
CC   cancerous condition and screening for or diagnosing cancerous conditions.
CC   The cancer associated antigens can be used as an immunogenic or vaccine
CC   composition with an adjuvant, e.g. a cytokine, a saponin, or granulocyte
CC   macrophage-colony stimulating factor (GM-CSF)
XX
SQ   Sequence 3283 BP; 945 A; 833 C; 779 G; 725 T; 0 U; 1 Other;

Query Match          92.1%;   Score 3143;   DB 3;   Length 3283;
Best Local Similarity  96.2%;
Matches 3283;   Conservative   0;   Mismatches   0;   Indels  129;   Gaps   1;
```

```
Qy    1  GGCAGCGGAGGAGGCGAGGAGGCGAGGAGGCCCGGGACCGGGCCGGGGAGCCGTCTCGGGG   60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  GGCAGCGGAGGAGGCGAGGAGGCGAGGAGGCCCGGGACCGGGCCGGGGAGCCGTCTCGGGG   60

Qy   61  AAGAGACGGATGATGAACAAGCTTTACATCGGGACAACCTGAGCCCCGTCACCGCCGAC  120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  AAGAGACGGATGATGAACAAGCTTTACATCGGGACAACCTGAGCCCCGTCACCGCCGAC  120

Qy  121  GACCTCCGGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGGCGGGACAGGTCCTGCGAAG  180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  GACCTCCGGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGGCGGGACAGGTCCTGCGAAG  180

Qy  181  TCCGGCTACGGCCTTCGTGGACTACCCCGGACTCGGGACAGTGGACAAGTCAACAGAC  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TCCGGCTACGGCCTTCGTGGACTACCCCGGACTCGGGACAGTGGACAAGTCAACAGAC  240

Qy  241  CTCTCGGGTAAAGTGAATTGCATGGGAAAATCATGGGAAGTTGATTGATTCGAGTCTCTAAA  300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  CTCTCGGGTAAAGTGAATTGCATGGGAAAATCATGGGAAGTTGATTGATTCGAGTCTCTAAA  300

Qy  301  AAGCTAAGGAGCAGGAGAAATTCAGATTCGAAACATCCCTCCTCCACCTGCGACTGGGAGGTG  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  AAGCTAAGGAGCAGGAGAAATTCAGATTCGAAACATCCCTCCTCCACCTGCGACTGGGAGGTG  360

Qy  361  TTGGATGGACTTTGGCTCAATATGGGACAGTGGGACAAGTGTGGACAAGTGAACAAGTCAACAGAC  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  TTGGATGGACTTTGGCTCAATATGGGACAGTGGGACAAGTGTGGACAAGTGAACAAGTCAACAGAC  420

Qy  421  ACAGAAACCGCGTTGCTCAAGTCACATATGTCAACGTGCTCAACAAGAGAGCAAAAATAGCCATG  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  ACAGAAACCGCGTTGCTCAAGTCACATATGTCAACGTGCTCAACAAGAGAGCAAAAATAGCCATG  480

Qy  481  GAGAAGCTAAGCGGCATCAGTTGAGAACTACTCCTTCCTTCAAGATTCCTACATCCGGAT  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  GAGAAGCTAAGCGGCATCAGTTGAGAACTACTCCTTCCTTCAAGATTCCTACATCCGGAT  540

Qy  541  GAGAGGTGAGCTCCCCTTCGCCCCGTCCCCTTCGCCCCGTCCCCAGCGTGGGGACCACTCTTCCCGG  600
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  GAAGAGGTGAGCTGAGCTCCCCTTCGCCCCGTCCCCAGCGTGGGGACCACTCTTCCCGG  600

Qy  601  GAGCAAGGCCACGCCCGCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTCCCGCTGCGG  660
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  GAGCAAGGCCACGCCCGCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTCCCGCTGCGG  660

Qy  661  ATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGAGAGGGGCTTGACCATAAAG  720
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661  ATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGAGAGGGGCTTGACCATAAAG  720

Qy  721  AACATCACTAAGCAGACCCAGTCCCGGTAGATATCCGGTAGATATCCATAGAAAGAGAACTCTGGAGCT  780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721  AACATCACTAAGCAGACCCAGTCCCGGTAGATATCCATAGAAAGAGAACTCTGGAGCT  780

Qy  781  GCAGGAAGCTGTCACCATCCATGTGCCACCCCAGAGGGGGACTTCTGAGCATGCGCATG  840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781  GCAGAGAAGCCTGTCACCATCCATGTGCCACCCCAGAGGGGGACTTCTGAAGCATGCGCATG  840

Qy  841  ATTCTTGAAATCATTGCAGAAAGAGAGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG  900
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841  ATTCTTGAAATCATTGCAGAAAGAGAGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG  900

Qy  901  AAAATCTTGGCACACAATGGCTTCGTTGGAAGACTGATTGGAAGAAGGAGGCAGAAATTTG  960
            |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db  901  AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAGAAGGAGGCAGAAATTTG  960

Qy  961  AAGAAAATTGAACATGAAACAGGGACCAAGATAACAAATCTCATCTTTGCAGGATTTGAGC 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961  AAGAAAATTGAACATGAAACAGGGACCAAGATAACAAATCTCATCTTTGCAGGATTTGAGC 1020

Qy 1021  ATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTTGTCCAGTGCT 1080
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021  ATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTTGTCCAGTGCT 1080

Qy 1081  GAGATAGAGATTATGAAGAAGCTGCGTGAGGAGGCCTTGAAAATGATATGCGGCTGTTAAC 1140
            |||||||||||||||||||||||||||| |||||||||||||| |||||||||||||||
Db 1081  GAGATAGAGATTATGAAGAAGCTGCGTGAGGAGGCCTTGAAAATGATATGCGGCTGTTAAC 1140

Qy 1141  CAACAAGCCAATCTGATCCGCAGGGTTGAACCTCGAGCCACTTGGCATCTTTCAACAGGA 1200
Db 1141  --------------------------------------------------------- 1140

Qy 1201  CTGTCCGTGTATCTCCACAGCAGGGGCCCGGGGACGCTCCCCCGTCCCCCTACCAC 1260
Db 1141  --------------------------------------------------------- 1140

Qy 1261  CCCTTCACTACCCACTCCGGATACTTCTCCGACAGGAGCAGGGCCATCGGCCATCGAGACC 1320
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141  ------ACCCACTCCGGATACTTCTCCGACAGGAGCAGGGCCATCGGCCATCGAGACC 1191

Qy 1321  TTCCCGCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTCTTCATCCCAACCCAG 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1192  TTCCCGCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTCTTCATCCCAACCCAG 1251

Qy 1381  GCTGTGGGCGCAGCTCTTTGGGGAAGAGGGGCACCATCAAACAGCTGGCGAGATTCGCC 1440
            |||||||||||        ||||||||||||||||||||||||||||||||||||||||
Db 1252  GCTGTGGGCGCCATCAGCTGGCGAGATTCGCC 1311

Qy 1441  GGAGCCTCTATCAAGATTGCCCTGCGGAAGGCCCAGCGAAAGGATGGTCATC 1500
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1312  GGAGCCTCTATCAAGATTGCCCTGCGGAAGGCCCAGCGAAAGGATGGTCATC 1371

Qy 1501  ATCACCGGGCACCGAAGCCAGTTCAGGCCCCAGGGACGGATCTTTGGGAAACTGAAA 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1372  ATCACCGGGCACCGAAGCCAGTTCAGGCCCCAGGGACGGATCTTTGGGAAACTGAAA 1431

Qy 1561  GAGGAGAAACTTCTTTAACCCAAAGAAGAGAGTGAGAGCTCGAAGCGGCATATCGAGTGCCC 1620
```

```
Db   1432 GAGGAAAACTTCTTTAACCCCAAAGAGAGAGTGAAGCTGAAGCCGCATATCAGATGTGCCC 1491
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1621 TCTTCCACAGTCTGGCGGGTGATTGGCAAGACCGTTGGCAAGAACGAACTGCAGAAC 1680
           ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   1492 TCTTCCACAGTCTGGCGGGTGATTGGCAAGACGTTGGCAAGAACGAACTGCAGAAC 1551

Qy   1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCAGATGAAAATGAGAGAAGTG 1740
           ||||||||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db   1552 TTAACCAGTGCAGAAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCAGATGAAAATGAGAGAAGTG 1611

Qy   1741 ATCGTCAGAATTATCGGGCACTTCTTGCTAGCCAGACTGCACGCGCAAGATCAGGGAA 1800
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1612 ATCGTCAGAATTATCGGGCACTTCTTGCTAGCCAGACTGCACGCGCAAGATCAGGGGA 1671

Qy   1801 ATTGTACAACAGGTGAGCGAGCAGCAGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGC 1860
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1672 ATTGTACAACAGGTGAGCGAGCAGCAGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGC 1731

Qy   1861 AGCAAGTGAGCTCCCACAGCACCAGCAAACACGGATGAATGTAGCCCTTCCAAAC 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1732 AGCAAGTGAGCTCCCACAGCACCAGCAAACACGGATGAATGTAGCCCTTCCAAAC 1791

Qy   1921 CTGACAGAATGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAG 1980
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1792 CTGACAGAATGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAG 1851

Qy   1981 GAATGAGAAGTCTGCGGAGGCGGGCCAGGCTCTGCCAGGGCTCTGCCAGGGGC 2040
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1852 GAATGAGAAGTCTGCGGAGGCGGGCCAGGCTCTGCCAGGGCTCTGCCAGGGGC 1911

Qy   2041 CGAGGAGGGGGCGGGGGCCAGGTTCAGCCAGGTTGCCAGAACCACCACCGAGCCCCGCC 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1912 CGAGGAGGGGGCGGGGGCCAGGTTCAGCCAGGTTGCCAGAACCACCACCGAGCCCCGCC 1971

Qy   2101 CCCCAGGGCTTCTGCGAGGCTTCAGCCATCCACACTTCACCATCCACTCGTCTCCTGAA 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1972 CCCCAGGGCTTCTGCGAGGCTTCAGCCATCCACACTTCACCATCCACTCGTCTCCTGAA 2031

Qy   2161 CTCCCACGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2032 CTCCCACGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2091

Qy   2221 AAAATGCACACCCTTTTCGTGTGGCAAATCGTCTCTGTACATGTGTATACATATTAGAAA 2280
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2092 AAAATGCACACCCTTTTCGTGTGGCAAATCGTCTCTGTACATGTGTATACATATTAGAAA 2151

Qy   2281 GGGAAGATGTTAAGATATGTGGCCTGTGTGGGTTACACAGGGTGCCTGCCTCACAGGGTGC 2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2152 GGGAAGATGTTAAGATATGTGGCCTGTGTGGGTTACACAGGGTGCCTGCCTCACAGGGTGC 2211

Qy   2341 TTTAGAAATAATATCAAATAACTCAACTAACTCCAATTTTAATCAATTCAATTATTAATTTT 2400
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2212 TTTAGAAATAATATCAAATAACTCAACTAACTCCAATTTTAATCAATTCAATTATTAATTTT 2271

Qy   2401 TTTTTCTTTTTAAAGAGAAGAAGCAGCTTTCTCAGACTTTAAAGAATAAAGTCTTTGGGAG 2460
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2272 TTTTTCTTTTTAAAGAGAAGAAGCAGCTTTCTCAGACTTTAAAGAATAAAGTCTTTGGGAG 2331

Qy   2461 GTCTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGAGGGAAAT 2520
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2332 GTCTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGAGGGAAAT 2391

Qy   2521 CTCGTCGGAAGGACACTCACGCAGTTCTGGATCACCTGTATGTCAACAGAAGAGGATA 2580
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2392 CTCGTCGGAAGGACACTCACGCAGTTCTGGATCACCTGTATGTCAACAGAAGAGGATA 2451

Qy   2581 CCGTCTCCTTGAAGCAGGAACTCTGTCACTCGTCTAGTCTCAATGCCTGTCATCACCCATT 2640
           ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   2452 CCGTCTCCTTGAAGCAGGAACTCTGTCACTCGTCTAGTCTCATGCCTGTCATCACCCATT 2511

Qy   2641 TCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCT 2700
           |||||||||||||||||||
```

```
Db   2512 TCTCTTTGCTTCACAGGTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCT 2571

Qy   2701 CTCTCTGTTTATCTCCCATCCCTTCCCTCCCATCTCCATTCTTCTTTTGAA 2760
           |||||||||||| ||||||||||||||||||||||||||||||||| ||||
Db   2572 CTCTCGTTTGTTATCTTCTCTCCCTCTCCCCTTCTCCATCTCCATTCTTTTGAA 2631

Qy   2761 TTTCCTCATCCCTCCATCTCAATCCCGTATCTACGACCCCCCCCCCCCAGGCAAAGCA 2820
           |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db   2632 TTTCCTCATCCCTCCATCTCAATCCCGTATCTACGCACCCCCCCCCAGGCAAAGCA 2691

Qy   2821 GTGCTCTGAGTATCACATCACAAAAGGAACAAAAGCGAAACACACAAAACCAGCCTCAA 2880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2692 GTGCTCTGAGTATCACATCACAAAAGGAACAAAAGCGAAACACACAAAACCAGCCTCAA 2751

Qy   2881 CTTACACTTGGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAG 2940
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2752 CTTACACTTGGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAG 2811

Qy   2941 AGGAAAACAGGAACCCACCAACCAACCAATCAAAGAAAAATTCCACAATG 3000
           |||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   2812 AGGAAAACAGGAACCCACCAACCAACCAATCAACCAAAGAAAGAAAAATTCCACAATG 2871

Qy   3001 AAAGAAATGTATTTTGTCTTTTGTGTTATAAGCCATCAATATTCAGCAAAATG 3060
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2872 AAAGAAATGTATTTTGTCTTTTGTGTTATAAGCCATCAATATTCAGCAAAATG 2931

Qy   3061 ATTCCTTTCTTTAAAAAAAAATGTGGAGAAAGTAGAAATTTACCAAGGTTGTTGGCC 3120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2932 ATTCCTTTCTTTAAAAAAAAATGTGGAGAAAGTAGAAATTTACCAAGGTTGTTGGCC 2991

Qy   3121 CAGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAACACAGAGAAGCTACCTCAG 3180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2992 CAGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAACACAGAGAAGCTACCTCAG 3051

Qy   3181 GTGTTTTACCTCAGCACCTTGCTCTGTGTTTCCCTGGATCTCTCCTGATAG 3240
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3052 GTGTTTTACCTCAGCACCTTGCTCTGTGTTTCCCTGGATCTCTCCTGATAG 3111

Qy   3241 TTGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGCATATCAACT 3300
           |||||||||||||||||||| |||||||||||||||||||||||||||| ||||||||
Db   3112 TTGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACT 3171

Qy   3301 GCCAGCCTGGAGAAGGTGACAGTGCAAGTGTCCAAGTGTGCAACGTGTCTCTTCCGCT 3360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3172 GCCAGCCTGGAGAAGGTGACAGTGCAAGTGTCCAAGTGTGCAACGTGTCTCTTCCGCT 3231

Qy   3361 AGCCAAGAACCNATATGGCCTTCTTCTTTGGACAAACCTTGAAAATGTTTATTT 3412
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3232 AGCCAAGAACCNATATGGCCTTCTTCTTTGGACAAACCTTGAAAATGTTTATTT 3283
```

```
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG06795.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 6786; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3667 BP; 1081 A; 882 C; 846 G; 858 T; 0 U; 0 Other;

Query Match             91.3%;  Score 3114.8;  DB 5;  Length 3667;
Best Local Similarity   96.1%;  Pred. No. 0;
Matches 3277;  Conservative    0;  Mismatches    3;  Indels  131;  Gaps   3;

Qy      3 CAGCGGAGGAGGCGAGGAGGCCGGGTACCGGCCGGGGGAGCCGGGGCTCCGGGGAA      62
          ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    369 CAACGGAGGAGGCGGAGGAGGCCGCCGGGTACCGGCCGGGGAGCCGGGGCTCCGGGGAA   428

Qy     63 GAGACCGATGATGAACAAGCTTTACATCGGGAACTTGAGCCCCGTCACCGCCGACGA     122
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    429 GAGACCGATGATGAACAAGCTTTACATCGGGAACTTGAGCCCCGTCACCGCCGACGA    488

Qy    123 CCTCCGGCCAGCTCTTTGGGGACAGGAGCTGCCCCTGGCCCCTGGCCGCTGCTGAAGTC    182
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    489 CCTCCGGCCAGCTCTTTGGGGACAGGAGCTGCCCCTGGCCCCTGGCCGCTGCTGAAGTC   548

Qy    183 CGGCTACGCCCTTCGTGGACTACCCCGACCAGAACTGGGCCATCGGCGCGCCATCGAGAC    242
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    549 CGGCTACGCCCTTCGTGGACTACCCCGACCAGAACTGGGCCATCGGCGCGCCATCGAGAC   608

Qy    243 CTTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAA   302
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    609 CTCCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAA   668

Qy    303 GCTAAGGAGCAGGAAAATTCAGATTCAGAATTCAGAAACATCGAAACATCCTCACCTGCAGTGGAGTGTT    362
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    669 GCTAAGGAGCAGGAAAATTCAGATTCAGAATTCAGAAACATCCTCACCTGCAGTGGAGTGTT   728

Qy    363 GGATGGACTTTTGGCTCAATATGGGACAGTGGACAGTGAACAAGTCAACACAGACAC    422
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    729 GGATGGACTTTTGGCTCAATATGGGACAGTGGACAGTGAACAAGTCAACACAGACAC   788

Qy    423 AGAAACCGCCGTTGTCAACGTCACATATGCAACAAGAGAAGAAGCAAAAATAGCCATGGA    482
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    789 AGAAACCGCCGTTGTCAACGTCACATATGCAACAAGAGAAGAAGCAAAAATAGCCATGA   848

Qy    483 GAAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGATGA    542
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    849 GAAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGATGA   908

Qy    543 AGAGGTGAGCTCCCCTTCGCCCCTCAGCGAGCCCAGCGGCGGGGACCCACTCTTCCCGGGA    602
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    909 AGAGGTGAGTCCCCTTCGCCCCTCAGCGAGCCCAGCGGCGGGGACCCACTCTTCCCGGGA   968

Qy    603 GCAAGGCCACGCCCCTGGGGGCACTTCTCAGCGAGCACAGATTGATTCCCGCTGCGGAT    662
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    969 GCAAGGCCACGCCCCTGGGGGCACTTCTCAGCGAGCACAGATTGATTCCCGCTGCGGAT  1028

Qy    663 CCTGGTCCCCACCAGTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAA    722
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1029 CCTGGTCCCCACCAGTTGTTGGTGCCATCATCATGCGGAAAGGAGGGCTTGACCATAAAGAA  1088

Qy    723 CATCACTAAGCAGACCCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGGAGCTGC    782
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1089 CATCACTAAGCAGACCCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGGAGCTGC  1148

Qy    783 AGAGAAGCCGTCACCATCATGCCACCCAGAGGGGACTTCTGAAGCATGCCGCATGAT    842
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1149 AGAGAAGCCGTCACCATCATGCCACCCAGAGGGGACTTCTGAAGCATGCCGCATGAT  1208

Qy    843 TCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAA    902
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1209 TCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAA  1268

Qy    903 AATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAGAAAGGCAGAAATTTGAA    962
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1269 AATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAGAAAGGCAGAAATTTGAA  1328

Qy    963 GAAAATTGAACATGAAACAGGACCAAGAAACAATCTCATCTTTGCAGGATTTGAGCAT   1022
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1329 GAAAATTGAACATGAAACAGGACCAAGAAACAATCTCATCTTTGCAGGATTTGAGCAT  1388

Qy   1023 ATACAACCCGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTTGCCAGTGCTGA   1082
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1389 ATACAACCCGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTTGCCAGTGCTGA  1448

Qy   1083 GATAGAGATTATGAGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGCTGCTGTTAACCA   1142
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1449 GATAGAGATTATGAGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGCTGCTGTTAAC--  1506

Qy   1143 ACAAGCCAATCGATCCCAGGGTTGAACCTCAGCGGCACTTGAACATCTTTTCAACAGGAGACT   1202
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1507 ------------------------------------------------------------   1506

Qy   1203 GTCCGTGCTATCCTCCACCAGCAGGGCCCCCGCCCCGCTGCCCCTCACCACCC   1262
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1507 ------------------------------------------------------------   1506

Qy   1263 CTTCACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTGGCCCGTT   1322
          ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   1507 -----ACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTGGCCCGTT  1559

Qy   1323 CCCGCATCATCACTCTATCCAGCAGAGATTGTGAATCTCTTCATCCCAACCCAGGC   1382
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1560 CCCGCATCATCACTCTTATCCAGCAGAGGAGATTGTGAATCTCTTCATCCCAACCCAGGC  1619

Qy   1383 TGTGGGGCGCCATCATCGGGAAGAGAGGGGCACACATCAAACAGCTGGCGAGATTCGCCGG   1442
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1620 TGTGGGGCGCCATCATCGGGAAGAGAGGGGGCACACATCAAACAGCTGGCGAGATTCGCCGG  1679

Qy   1443 AGCCTCTATCAAGATTGCCCCTGCCCGGAAGGCCCAGCGTCAGCGGAAGGAGATGGTCATCAT   1502
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1680 AGCCTCTATCAAGATTGCCCCTGCCCGGAAGGCCCAGCGTCAGCGGAAGGAGATGGTCATCAT  1739

Qy   1503 CACCGGCGCACCGGAAGCCCAGTTCAAGGCCGATCTTTGGGAAAACTGAAGAGAAGAAAGA   1562
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    ...
```

```
Db   1740 CACCGGGCCACCGGAGGCCCAGTTCAAGGCCGAGGGACCGGATCTTGGGAAACTGAAAGA 1799
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1563 GGAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGGAAGCGCATATCAGAGTGCCCTC 1622
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1800 GGAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGGAAGCGCATATCAGAGTGCCCTC 1859

Qy   1623 TTCCACAGCTGCGCCGGGTGATTGGCAAGACCGTGAACGACTGCGACGAACTT 1682
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1860 TTCCACAGCTGCGCCGGGTGATTGGCAAGACCGTGAACGACTGCGACGAACTT 1919

Qy   1683 AACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAGACGCCAGATGAAAATGAGGAAGTGAT 1742
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1920 AACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTGAT 1979

Qy   1743 CGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAAT 1802
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1980 CGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAAT 2039

Qy   1803 TGTACAACAGTGAAGCAGCAGCAGAGAATACCCTCAGGGAGTCGCCTCACAGCGCAG 1862
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2040 TGTACAACAGTGAAGCAGCAGCAGAGAATACCCTCAGGGAGTCGCCTCACAGCGCAG 2099

Qy   1863 CAAGTGAGGCTCCCACAGGCCACCAGCAAACACGATGAATGTAGCCCTTCCAACACCT 1922
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2100 CAAGTGAGGCTCCCACAGGCCACCAGCAAACACGATGAATGTAGCCCTTCCAACACCT 2159

Qy   1923 GACAGAATGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAGACCATCTGAGA 1982
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2160 GACAGAATGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAGACCATCTGAGA 2219

Qy   1983 ATGAGAAGTCTGCGGAGGGCGGGGGAAGTCAGCCAGGTTCGCCAGAACCACCAGGGCCG 2042
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2220 ATGAGAAGTCTGCGGAGGGCGGGGGAAGTCAGCCAGGTTCGCCAGAACCACCAGGGGCG 2279

Qy   2043 AGGGAGGGGCGGGAAGTCAGCCAGGTTCGCCAGAACCACCAGGGCCCGCTCCCGCCCC 2102
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2280 AGGGAGGGGCGGGAAGTCAGCCAGGTTCGCCAGAACCACCAGGGCCCGCTCCCGCCCC 2339

Qy   2103 CCAGGGCTTCTCGCAGGCTTCAGCCATCCACCATCCACTTCACCACTTAGTGAATCTCCTGAACT 2162
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2340 CCAGGGCTTCTCGCAGGCTTCAGCCATCCACCATCCACTTCACCACTTAGTGAATCTCCTGAACT 2399

Qy   2163 CCCACAGCGCTATCCCTTTTCTGTGGCAAATCGTCTCTGTACATAGTGTTCAAAGCCAAGCAA 2222
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2400 CCCACAGCGCTATCCCTTTTCTGTGGCAAATCGTCTCTGTACATAGTGTTCAAAGCCAAGCAA 2459

Qy   2223 AATGCACACCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATAGTGTTAGAAAGG 2282
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2460 AATGCACACCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATAGTGTTAGAAAGG 2519

Qy   2283 GAAAATGTTAAGATGTTAAGATATGTGGCCTGTGGGGTTACACAGGGGTTACACAGGTGTAATAATATTT 2342
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2520 GAAGATGTTAAGATGTTAAGATATGTGGCCTGTGGGGTTACACAGGTGCCTGCAGGCGGTAATAATATTT 2579

Qy   2343 TAGAAATAATATATATCAAATAACTCAACTACTCAATTTTAATCAATTATTAATTTTTT 2402
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2580 TAGAAATAATATATATCAAATAACTCAACTACTCAATTTTAATCAATTATTAATTTTTT 2639

Qy   2403 TTTCTTTTTAAAGAAGAAAGCAGAGGCGCTTTGAGGCAGGAATAAAGTCTTTTGGGAGGT 2462
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2640 TTTCTTTTTAAAGAAGAAAGCAGAGGCGCTTTGAGGCAGGAATAAAGTCTTTTGGGAGGT 2699

Qy   2463 CTCACGGTGTGAGAGGACGAGCTTTGAGGCCACCCGACACAAATTCACCCGAGGGGAAATCT 2522
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2700 CTCACGGTGTGAGAGGACGAGCTTTGAGGCCACCCGACACAAATTCACCCGAGGGGAAATCT 2759

Qy   2523 CGTCGGAAGGACACTCACGGCAGTTCTGGGATCACCTGGTGTGCCTCAACAGGAGGGATACC 2582
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2760 CGTCGGAAGGACACTCACGGCAGTTCTGGGATCACCTGGTGTGCCTCAACAGGAGGGATACC 2819

Qy   2583 GTCTCCTTGAAGAGGAAACTCTGGCTCACTCTCCTCAGCTCTGCTGCTCTAGCCACCACCATTTC 2642
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2820 GTCTCCTTGAAGAGGAAACTCTGGCTCACTCTCCTCAGCTCTGCTGCTCTAGCCACCACCATTTC 2879
```

```
Qy   2643 TCTTTGCTTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCGTCTCT 2702
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2880 TCTTTGCTTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCGTCTCT 2939

Qy   2703 CTCTGTTTATCTCCCCTCCCCTCCCCTTCTTCTCCATCTCCATTCTCTTTGAATT 2762
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2940 CTCTGTTTATCTCCCCTCCCCTCCCCTTCTTCTCCATCTCCATTCTCTTTGAATT 2999

Qy   2763 TCCTCATCCCTCCTCAATCCCGTATCTAGGCACCCCCCCCCCCCAGGCAAAGCAGT 2822
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3000 TCCTCATCCCTCCTCAATCCCGTATCCCAATCCCGTATCTACGCA-CCCCCCCCCCCAGGCAAAGCAGT 3058

Qy   2823 GCTCGAGTATCATCACAAGAACAAAAGGAACAAAAGCGAACACACAAACCAGCCTCAACT 2882
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3059 GCTCGAGTATCACATCACAAGAACAAAAGGAACAAAAGCGAAACACACACAAACCAGCCTCAACT 3118

Qy   2883 TACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAG 2942
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3119 TACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAG 3178

Qy   2943 GAAAACAGGAACCCACCACCAATCAACCAACAAAGAAAAATTCCACAATGAA 3002
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3179 GAAAACAGGAACCCACCACCAATCAACCAACAAACAAGAAAAAATTCCACAATGAA 3238

Qy   3003 AGAATGTATTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGAT 3062
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3239 AGAATGTATTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGAT 3298

Qy   3063 TCCTTTCTTT-AAAAAAAAAAATGTGGAGGAAAGTAGAAATTACCAAGGTTGTTGGCCC 3121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3299 TCCTTTCTTTAAAAAAAAAAAAATGTGGAGGAAAGTAGAAATTACCAAGGTTGTTGGCCC 3358

Qy   3122 AGGGCGTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAGAAGAAGCTACCTCAGG 3181
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3359 AGGGCGTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAGAAGAAGCTACCTCAGG 3418

Qy   3182 TGTTTTTACCTCAGCACCTTGCTCTTGTGTTCCCTTAGAGATTTTGTAAAGCTGATAGT 3241
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3419 TGTTTTTACCTCAGCACCTTGCTCTTGTGTTCCCTTAGAGATTTTGTAAAGCTGATAGT 3478

Qy   3242 TGGAGCATTTTTTATTTTTAATAAAAATGAGTTGGAAAAAAATAAGAATATCAACTG 3301
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3479 TGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGAATATCAACTG 3538

Qy   3302 CCAGCCTGGAGAGGTGACAGTCCAAGTGCAACAGCGTGTCGAATTGTCTTCCGCTA 3361
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3539 CCAGCCTGGAGAGGTGACAGTCCAAGTGCGACAGCGTGTCGAATTGTCTTCCGCTA 3598

Qy   3362 GCCAAGAACCNATATGGCCTTCTTTTGGACAAATCCTTGAAAATGTTTATTT 3412
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3599 GCCAAGAACCTATATGGCCTTCTTTTTGGACAAACCTTGAAAATGTTTATTT 3649
```

```
RESULT 6
ABS76442
ID   ABS76442   standard; cDNA; 3667 BP.
XX
AC   ABS76442;
XX
DT   11-DEC-2002   (first entry)
XX
DE   cDNA encoding human ovarian cancer marker M452.
XX
KW   Human; ovarian cancer; marker; cancer; familial history; brain disorder;
KW   central nervous system disorder; bacterial meningitis; viral meningitis;
KW   Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus;
KW   brain herniation; inflammation; encephalitis; testicular disorder;
KW   nontuberculous granulomatous orchitis; connective tissue disorder;
KW   heart disorder; ischaemic heart disease; atherosclerosis; neoplasm;
KW   histological type; carcinogenic; ovarian cancer marker; gene; ss.
XX
OS   Homo sapiens.
```

```
XX
PN   WO200271928-A2.
XX
PD   19-SEP-2002.
XX
PF   14-MAR-2002; 2002WO-US007826.
XX
PR   14-MAR-2001; 2001US-0276025P.
PR   14-MAR-2001; 2001US-0276026P.
PR   10-AUG-2001; 2001US-0311732P.
PR   19-SEP-2001; 2001US-0323580P.
PR   26-SEP-2001; 2001US-0324967P.
PR   26-SEP-2001; 2001US-0325102P.
PR   26-SEP-2001; 2001US-0325149P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Monahan JE, Gannavarapu M, Hoersch S, Kamatkar S, Kovatis SG;
PI   Meyers RE, Morrisey MP, Olandt PJ, Sen A, Vieby PO, Mills GB;
PI   Bast RC, Lu K, Schmandt RE, Zhao X, Glatt K;
XX
DR   WPI; 2002-723277/78.
DR   P-PSDB; ABG96346.
XX
PT   Assessing whether a patient is afflicted with ovarian cancer, useful in
PT   assessing the stage or progression of the disease, comprises comparing
PT   the expression level of a cancer marker in a sample from a patient and
PT   from a non cancer patient.
XX
PS   Disclosure; Page 262-263; 481pp; English.
XX
CC   The present invention relates to a new method for assessing whether a
CC   patient is afflicted with ovarian cancer. The method involves comparing
CC   the expression level of a marker in a patient sample and the normal level
CC   of expression of the marker in a control non-ovarian cancer sample, where
CC   the marker is selected from 363 cancer markers described in the
CC   specification. The method of the invention is useful in diagnosing or
CC   characterising cancer, in detecting the presence of cancer as early as
CC   possible, and the recurrence of ovarian cancer. The method may also be of
CC   particular use with patients having an enhanced risk of developing
CC   ovarian cancer (e.g. patients having a familial history of ovarian
CC   cancer). The cancer markers may be used in the management and treatment
CC   of e.g. brain and central nervous system disorders (e.g. bacterial and
CC   viral meningitis, Alzheimer's disease or Parkinson's disease), brain
CC   disorders (e.g. cerebral oedema, hydrocephalus or brain herniations),
CC   inflammations (e.g. bacterial or viral meningitis or encephalitis),
CC   testicular disorders (e.g. nontuberculus granulomatous orchitis),
CC   connective tissue disorders, or heart disorders (e.g. ischaemic heart
CC   disease or atherosclerosis). The compositions and methods may also be
CC   used in assessing the histological type of neoplasm associated with
CC   ovarian cancer, monitoring the progression of ovarian cancer, determining
CC   whether ovarian cancer has metastasized or is likely to metastasize,
CC   selecting a composition for inhibiting ovarian cancer, assessing the
CC   ovarian carcinogenic potential of a compound, or inhibiting ovarian
CC   cancer or at risk of developing ovarian cancer. The present nucleic acid
CC   sequence encodes one of the ovarian cancer markers described in the
CC   invention
XX
SQ   Sequence 3667 BP; 1081 A; 882 C; 846 G; 858 T; 0 U; 0 Other;

Query Match     91.3%;  Score 3114.8;  DB 6;  Length 3667;
Best Local Similarity   96.1%;  Pred. No. 0;
Matches 3277;  Conservative    0; Mismatches   3; Indels 131; Gaps   3;
```

```
Qy      3 CAGCGGAGGAGGAGGCGGAGAGCGCCGGGTACCGGCCGCGGAGGCCGCCGGGTCTCGGGGAA  62
          ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db    369 CAACGGAGGAGGAGGCGGAGAGCGCCGGGTACCGGCCGCGGAGGCCGCCGGGTCTCGGGGAA  428

Qy     63 GAGACGGATGATGAACAAGCTTTACATCGGGAACCTTGAGCCCGTCACCGCCGACGA  122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    429 GAGACGGATGATGAACAAGCTTTACATCGGGAACCTTGAGCCCGTCACCGCCGACGA  488

Qy    123 CCTCCCGGCCAGCTCTTTGGGGACCAGGAAGCTGCCCCTGGGCGGGACAGGTCTCTGCTGAAGTC  182
```

Db   1506   --   1506

Qy   1263   CTTCACTACCCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTGGCCCGTT   1322
            |||||||||||||||
Db   1507   ------ACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTGGCCCGTT   1559

Qy   1323   CCCGCATCACTCATCTTATCCAGAGCAGGAGATTGTGAATCTCTTCATCCCAACCCAGGC   1382
            |||||||||||||||
Db   1560   CCCGCATCACTCATCTTATCCAGAGCAGGAGATTGTGAATCTCTTCATCCCAACCCAGGC   1619

Qy   1383   TGTGGGCGCCATCATCGGGAAGAGGGGGCACACATCAAACAGTTGGCGAGATTCGCCCG   1442
            |||||||||||||||
Db   1620   TGTGGGCGCCATCATCGGGAAGAGGGGGCACACATCAAACAGTTGGCGAGATTCGCCCG   1679

Qy   1443   AGCCTCTATCAAGATTGCCCCTGCCGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCAT   1502
            |||||||||||||||
Db   1680   AGCCTCTATCAAGATTGCCCCTGCCGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCAT   1739

Qy   1503   CACCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGACGATCTTTGGGAAACTGAAAGA   1562
            |||||||||||||||
Db   1740   CACCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGACGATCTTTGGGAAACTGAAAGA   1799

Qy   1563   GGAAAACTTCTTTAACCCCAAAGAGAAGAAGTGAAGCGCATATCAGAGTGCCCTC   1622
            |||||||||||||||
Db   1800   GGAAAACTTCTTTAACCCCAAAGAGAAGAAGTGAAGCGCATATCAGAGTGCCCTC   1859

Qy   1623   TTCCACAGTCTGCCCGGTGATTGGGCAAAGGTGGCAAGACCGTGAACGAACTGCAGAACTT   1682
            |||||||||||||||
Db   1860   TTCCACAGTCTGCCCGGTGATTGGGCAAAGGTGGCAAGACCGTGAACGAACTGCAGAACTT   1919

Qy   1683   AACCAGTGCAGAAGTCATCGTCCTCGTGACCAAATGAAAATGAGGAAGTGAT   1742
            |||||||||||||||
Db   1920   AACCAGTGCAGAAGTCATCGTCCTCGTGACCAAATGAAAATGAGGAAGTGAT   1979

Qy   1743   CGTCAGAAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAAT   1802
            |||||||||||||||
Db   1980   CGTCAGAAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAAT   2039

Qy   1803   TGTACAACAGTGAAGCGAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAG   1862
            |||||||||||||||
Db   2040   TGTACAACAGTGAAGCGAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAG   2099

Qy   1863   CAAGTGAGGCTCCCACAGGCCACCAGCAAACAACGGATGAATGTAGCCCTTCCAACACCT   1922
            |||||||||||||||
Db   2100   CAAGTGAGGCTCCCACAGGCCACCAGCAAACAACGGATGAATGTAGCCCTTCCAACACCT   2159

Qy   1923   GACAGAGAATGAGAGACCAAACGCAGCCAGCAGCTGGGAGCAAACCACCATCTGAGGA   1982
            |||||||||||||||
Db   2160   GACAGAGAATGAGAGACCAAACGCAGCCAGCAGCTGGGAGCAAACCACCATCTGAGGA   2219

Qy   1983   AGGAGGGGCGGGGAAGGTCAGCCAGGTTGCCAGAACCACCACCGAGCCCTCCCGCCCC   2042
            |||||||||||||||
Db   2220   AGGAGGGGCGGGGAAGGTCAGCCAGGTTGCCAGAACCACCACCGAGCCCTCCCGCCCC   2279

Qy   2043   CCAGGGCTTCTGCAGGCTTCGCAGGCCATCCACTTCACCATCGGATCTCTCCTGAACT   2102
            |||||||||||||||
Db   2280   CCAGGGCTTCTGCAGGCTTCGCAGGCCATCCACTTCACCATCGGATCTCTCCTGAACT   2339

Qy   2103   CCCACGACCGTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAA   2162
            |||||||||||||||
Db   2340   CCCACGACCGTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAA   2399

Qy   2163   AATGCACACACCCTTTTTCTGTGCAAATCGTCTCTGTACATGTGTACATATTAGAAGG   2222
            |||||||||||||||
Db   2400   AATGCACACACCCTTTTTCTGTGCAAATCGTCTCTGTACATGTGTACATATTAGAAGG   2459

Qy   2223   GAAGATGTTAAGATGTGTGGGTTACACAGGGTGCCTGCAGCGTAATATATTT   2282
            |||||||||||||||
Db   2460   GAAGATGTTAAGATGTGTGGGTTACACAGGGTGCCTGCAGCGTAATATATTT   2519

Qy   2283   TAGAAATATATATCAAATAACTAGACTAACTCCAATTTTTAATCAATTATTAATTTTTT   2402
            |||||||||||||||
Db   2520   TAGAAATATATATCAAATAACTAACTCAACTAACTCCAATTTTTAATCAATTATTAATTTTTT   2639

Qy   2343   TTTCTTTTTAAAGAGAAGACAGCCGCTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGT   2462
            |||||||||||||||
Db   2580   TTTCTTTTTAAAGAGAAGACAGCCGCTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGT   2699

Qy   2463   CTCACGGTGTAGAGAGAGCTTTGAGGCGCCACCCGCACAAAAATTCACCCAGAGGGAAATCT   2522
            |||||||||||||||
Db   2700   CTCACGGTGTAGAGAGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGAGGGAAATCT   2759

Qy   2523   CGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAAGGGATACC   2582
            |||||||||||||||
Db   2760   CGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAAGGGGATACC   2819

Qy   2583   GTCTCCTTGAAGAGGAAACTCTGTCACTCCTCAGCCTGTCATACACCCATTTC   2642
            |||||||||||||||
Db   2820   GTCTCCTTGAAGAGGAAACTCTGTCACTCCTCAGCCTGTCATACACCCATTTC   2879

Qy   2643   TCTTTGCTTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCT   2702
            |||||||||||||||
Db   2880   TCTTTGCTTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCT   2939

Qy   2703   CTCTGTTTATCTCTCCCCTCCCCCCCCCCCCCCCCCCCCCCCCAGGCAAAGCAGT   2762
            |||||||||||||||
Db   2940   CTCTGTTTATCTCTCCCCTCCCTCCCCCTTCTCTCCATTCTTTGAATT   2999

Qy   2763   TCCTCATCCCTCCATCTCAATCCGTATCTACGCACCCCCCCCCCCCCCAGGCAAAGCAGT   2822
            |||||||||||||||
Db   3000   TCCTCATCCCTCCATCTCAATCCGTATCTACGCA-CCCCCCCCCCCCAGGCAAAGCAGT   3058

Qy   2823   GCTCTGAGTATCACATCACACACAAAAGCAAAAGCGAAACACACAAACCAGCCTCAACT   2882
            |||||||||||||||
Db   3059   GCTCTGAGTATCACATCACACACAAAAGCAAAAGCGAAACACACAAACCAGCCTCAACT   3118

Qy   2883   TACACTTGGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAG   2942
            |||||||||||||||
Db   3119   TACACTTGGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAG   3178

Qy   2943   GAAAACAGGAACCCACCAACCAACCAATCAACAAAGAAAAATTCCACAATGAA   3002
            |||||||||||||||
Db   3179   GAAAACAGGAACCCACCAACCAACCAATCAACAAAGAAAAATTCCACAATGAA   3238

Qy   3003   AGAATGTATTTTGTCTTTTGGTGTATAAGCCATCAATATTCAGCAAAATGAT   3062
            |||||||||||||||
Db   3239   AGAATGTATTTTGTCTTTTGGTGTATAAGCCATCAATATTCAGCAAAATGAT   3298

Qy   3063   TCCTTTCTT-AAAAAAAAAAATGAGTTGGAAAAAAATAAGATATCAACTG   3121
            |||||||||||||||
Db   3299   TCCTTTCTTTAAAAAAAAAAAAATGAGTTGGAAAAAAATAAGATATCAACTG   3358

Qy   3122   AGGGCGGTAAATTCACAGATTTTTTTAACGAGAAAAACACACAGAGAAGCTACCTCAGG   3181
            |||||||||||||||
Db   3359   AGGGCGGTAAATTCACAGATTTTTTTAACGAGAAAAACACACAGAGAAGCTACCTCAGG   3418

Qy   3182   TGTTTTTACCTCAGCACCTTGCTCTCTGTGTTTCCCTTAGACATTTGTAAAGCTGATAGT   3241
            |||||||||||||||
Db   3419   TGTTTTTACCTCAGCACCTTGATCTCTGTGTTTCCCTTAGACATTTGTAAAGCTGATAGT   3478

Qy   3242   TGGAGCATTTTTTTATTTTTTTAATAAAAAATGAGTTGGAAAAAAATAAGATATCAACTG   3301
            |||||||||||||||
Db   3479   TGGAGCATTTTTTTATTTTTTTAATAAAAAATGAGTTGGAAAAAAATAAGATATCAACTG   3538

Qy   3302   CCAGCCTGGAGAAGGTGACAGTCCAAGTGTCGCAACAGCTGTTCTTCTTCCGCTA   3361
            |||||||||||||||
Db   3539   CCAGCCTGGAGAAGGTGACAGTCCAAGTGTCGCAACAGCTGTTCTTCTTCCGCTA   3598

Qy   3362   GCCAAGAACCNNATATGGCCTTCTTTTTGGACAAACCTTGAAAATGTTTATTT   3412
            |||||||||||||||
Db   3599   GCCAAGAACCTATATGGCCTTCTTTTTGGACAAACCTTGAAAATGTTTATTT   3649

```
RESULT 7
AAS26153
AAS26153
ID   AAS26153 standard; cDNA; 2290 BP.
XX
AC   AAS26153;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human cDNA encoding a novel secreted protein, Seq ID 332.
XX
KW   Human; immunosuppressive; antiarthritic; ss; antirheumatic; cytostatic;
KW   cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW   antibacterial; virucide; fungicide; opthalmalogical; vulnerary;
KW   secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; skin ageing; food additive;
KW   preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001341.
XX
PR   31-JAN-2000; 2000US-0179065P.       PR   08-SEP-2000; 2000US-0231242P.
PR   04-FEB-2000; 2000US-0180628P.       PR   08-SEP-2000; 2000US-0231243P.
PR   24-FEB-2000; 2000US-0184664P.       PR   08-SEP-2000; 2000US-0231244P.
PR   02-MAR-2000; 2000US-0186350P.       PR   08-SEP-2000; 2000US-0231413P.
PR   16-MAR-2000; 2000US-0189874P.       PR   08-SEP-2000; 2000US-0231414P.
PR   17-MAR-2000; 2000US-0190076P.       PR   08-SEP-2000; 2000US-0232080P.
PR   18-APR-2000; 2000US-0198123P.       PR   12-SEP-2000; 2000US-0232081P.
PR   19-MAY-2000; 2000US-0205515P.       PR   14-SEP-2000; 2000US-0231968P.
PR   07-JUN-2000; 2000US-0209467P.       PR   14-SEP-2000; 2000US-0232397P.
PR   28-JUN-2000; 2000US-0214886P.       PR   14-SEP-2000; 2000US-0232398P.
PR   30-JUN-2000; 2000US-0215135P.       PR   14-SEP-2000; 2000US-0232399P.
PR   07-JUL-2000; 2000US-0216647P.       PR   14-SEP-2000; 2000US-0232400P.
PR   07-JUL-2000; 2000US-0216880P.       PR   14-SEP-2000; 2000US-0232401P.
PR   11-JUL-2000; 2000US-0217487P.       PR   14-SEP-2000; 2000US-0233063P.
PR   11-JUL-2000; 2000US-0217496P.       PR   14-SEP-2000; 2000US-0233064P.
PR   14-JUL-2000; 2000US-0218290P.       PR   14-SEP-2000; 2000US-0233065P.
PR   26-JUL-2000; 2000US-0220963P.       PR   21-SEP-2000; 2000US-0234223P.
PR   26-JUL-2000; 2000US-0220964P.       PR   21-SEP-2000; 2000US-0234274P.
PR   14-AUG-2000; 2000US-0224518P.       PR   25-SEP-2000; 2000US-0234937P.
PR   14-AUG-2000; 2000US-0224519P.       PR   25-SEP-2000; 2000US-0234998P.
PR   14-AUG-2000; 2000US-0225213P.       PR   26-SEP-2000; 2000US-0235484P.
PR   14-AUG-2000; 2000US-0225214P.       PR   27-SEP-2000; 2000US-0235834P.
PR   14-AUG-2000; 2000US-0225266P.       PR   27-SEP-2000; 2000US-0258365P.
PR   14-AUG-2000; 2000US-0225267P.       PR   29-SEP-2000; 2000US-0263227P.
PR   14-AUG-2000; 2000US-0225268P.       PR   29-SEP-2000; 2000US-0263267P.
PR   14-AUG-2000; 2000US-0225270P.       PR   29-SEP-2000; 2000US-0263368P.
PR   14-AUG-2000; 2000US-0225447P.       PR   29-SEP-2000; 2000US-0263369P.
PR   14-AUG-2000; 2000US-0225757P.       PR   29-SEP-2000; 2000US-0236370P.
PR   14-AUG-2000; 2000US-0225758P.       PR   02-OCT-2000; 2000US-0236802P.
PR   14-AUG-2000; 2000US-0225759P.       PR   02-OCT-2000; 2000US-0237037P.
PR   18-AUG-2000; 2000US-0226279P.       PR   02-OCT-2000; 2000US-0237038P.
PR   22-AUG-2000; 2000US-0226681P.       PR   02-OCT-2000; 2000US-0237039P.
PR   22-AUG-2000; 2000US-0226868P.       PR   02-OCT-2000; 2000US-0237040P.
PR   22-AUG-2000; 2000US-0227009P.       PR   13-OCT-2000; 2000US-0239935P.
PR   23-AUG-2000; 2000US-0227182P.       PR   13-OCT-2000; 2000US-0239937P.
PR   30-AUG-2000; 2000US-0228924P.       PR   20-OCT-2000; 2000US-0240960P.
PR   01-SEP-2000; 2000US-0229287P.       PR   20-OCT-2000; 2000US-0241221P.
PR   01-SEP-2000; 2000US-0229343P.       PR   20-OCT-2000; 2000US-0241785P.
PR   01-SEP-2000; 2000US-0229344P.       PR   20-OCT-2000; 2000US-0241786P.
PR   01-SEP-2000; 2000US-0229345P.       PR   20-OCT-2000; 2000US-0241787P.
PR   05-SEP-2000; 2000US-0229509P.       PR   20-OCT-2000; 2000US-0241808P.
PR   05-SEP-2000; 2000US-0229513P.       PR   20-OCT-2000; 2000US-0241809P.
PR   06-SEP-2000; 2000US-0230437P.       PR   20-OCT-2000; 2000US-0241826P.
PR   06-SEP-2000; 2000US-0230438P.       PR   01-NOV-2000; 2000US-0244617P.
                                         PR   08-NOV-2000; 2000US-0246474P.
                                         PR   08-NOV-2000; 2000US-0246475P.
                                         PR   08-NOV-2000; 2000US-0246476P.
                                         PR   08-NOV-2000; 2000US-0246477P.
                                         PR   08-NOV-2000; 2000US-0246478P.
                                         PR   08-NOV-2000; 2000US-0246523P.
                                         PR   08-NOV-2000; 2000US-0246524P.
                                         PR   08-NOV-2000; 2000US-0246525P.
                                         PR   08-NOV-2000; 2000US-0246526P.
                                         PR   08-NOV-2000; 2000US-0246527P.
                                         PR   08-NOV-2000; 2000US-0246528P.
                                         PR   08-NOV-2000; 2000US-0246532P.
                                         PR   08-NOV-2000; 2000US-0246609P.
                                         PR   08-NOV-2000; 2000US-0246610P.
                                         PR   08-NOV-2000; 2000US-0246611P.
                                         PR   08-NOV-2000; 2000US-0246613P.
                                         PR   17-NOV-2000; 2000US-0249207P.
                                         PR   17-NOV-2000; 2000US-0249208P.
                                         PR   17-NOV-2000; 2000US-0249209P.
                                         PR   17-NOV-2000; 2000US-0249210P.
                                         PR   17-NOV-2000; 2000US-0249211P.
                                         PR   17-NOV-2000; 2000US-0249212P.
                                         PR   17-NOV-2000; 2000US-0249213P.
                                         PR   17-NOV-2000; 2000US-0249214P.
                                         PR   17-NOV-2000; 2000US-0249215P.
                                         PR   17-NOV-2000; 2000US-0249216P.
                                         PR   17-NOV-2000; 2000US-0249217P.
                                         PR   17-NOV-2000; 2000US-0249218P.
                                         PR   17-NOV-2000; 2000US-0249244P.
```

```
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251556P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-488783/53.
DR   P-PSDB; AAU16166.
XX
PT   New nucleic acid molecules encoding 461 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives.
XX
PS   Claim 1; SEQ ID NO 332; 980pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Antibodies to the proteins can also be used in
CC   alleviating symptoms associated with the disorders and in diagnostic
CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC   and many other disorders listed in the specification. The polypeptides
CC   can also be used to aid wound healing and epithelial cell proliferation,
CC   to prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence encodes a novel secreted protein of the invention. Note: The
CC   sequence data for this patent did not form part of the printed
```

```
Query Match            58.3%;  Score 1989.8;  DB 4;  Length 2290;
Best Local Similarity  99.4%;  Pred. No. 0;
Matches  2028;  Conservative  0;  Mismatches  0;  Indels  8;  Gaps  3;

Qy   1376 CCCAGGCTGTGGCGCCATCATCGGGAAGAAGGGGGCACACATCAAACAGCTGGCGAGAT 1435
               ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25 CCCACGCGTCGGCGCCATCATCGGGAAGAAGGGGGCACACATCAAACAGCTGGCGAGAT 84

Qy   1436 TCGCGGAGCCTCTATCAAGATTGCCCCTGCGAAGGCCCAGCGAAGGATGG 1495
               ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     85 TCGCGGAGCCTCTATCAAGATTGCCCCTGCGAAGGCCCAGCGAAGGATGG 144

Qy   1496 TCATCATCACCGGCCACCGGAAGCCCAGTTCAAGGCCCAGGACGGATCTTTGGGAAAC 1555
               ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    145 TCATCATCACCGGCCACCGGGAGCCCAGTTCAAGGCCCAGGACGGATCTTTGGGAAAC 204
```

```
Qy   1556 TGAAAGAGAGAAACTTCTTTAACCCCAAAGAAGAAGAAGTGAGGCTGGAGGCGCATATCAGAG 1615
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    205 TGAAAGAGAGAAACTTCTTTAACCCCAAAGAAGAAGAAGTGAGGCTGGAGGCGCATATCAGAG 264

Qy   1616 TGCCCTCTTCCACAGCTGGCCGGGTGATTGGCAAGAGGTGGCAAGACCGTGAACGAACTGC 1675
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    265 TGCCCTCTTCCACAGCTGGCCGGGTGATTGGCAAGAGGTGGCAAGACCGTGAACGAACTGC 324

Qy   1676 AGAACTTAACCAGTGCAGAAGTCATCATCGGTCCTCGTGACCAAACGCCAGATGAAAATGAGG 1735
               ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    325 AGAACTTAACCAGTGCAGAAGTCATCGTCGCCTCGTGACCAAACGCCAGATGAAAATGAGG 384

Qy   1736 AAGTGATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCA 1795
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    385 AAGTGATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCA 444

Qy   1796 GGGAAATTGTACAACAGTGAAGCAGCAGACGCAGAAATACCCTCAGGGAGTCGCCTCAC 1855
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    445 GGGAAATTGTACAACAGTGAAGCAGCAGACGCAGAAATACCCTCAGGGAGTCGCCTCAC 504

Qy   1856 AGGCAGCAAGTGAGGCTCCCACAGGCACCAGCAACAACCGATGAATGTAGCCCTTCC 1915
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    505 AGGCAGCAAGTGAGGCTCCCACAGGCACCAGCAACAACCGATGAATGTAGCCCTTCC 564

Qy   1916 AACACCTGACAGAATGAGACCAGCCAGCCAGATCGGGAGCAAACCAAAGACCAT 1975
               ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    565 AACACCTGACAGAATGAGACCAGCCAGCCAGATCGGGAGCAAACCAAAGACCAT 624

Qy   1976 CTGAGGAATGAGAAGTCTGCGGAGGCGGCCAGGAGTCAGCCAGTTGCCAGAACCACCACCCCA 2035
               ||||||||||||||||||||||||||||||||||||||||||||||||||  |||||||||||
Db    625 CTGAGGAATGAGAAGTCTGCGGAGGCGGCCAGGAGTCAGCCAGTTGCCGGAGAACCCCA 684

Qy   2036 GGGGCCGAGGAGGGGGCCGGGAAGTCAGCTGTTGCCAGAACCACCGAGCCCCCTC 2095
               |||||||||||||||||||| |||||| |||||||||||||||||||||||| |||
Db    685 GGGGCCGAGGAGGGGGCCGGGGAAGTCAGCCAGTTGCCAGAACCACCGAGCCCCGCTC 744

Qy   2096 CCGCCCCCAGGCTTCTGCAGGCTTCAGCATCCACTTCACCATCCACTTCGGATCTCTC 2155
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    745 CCGCCCCCAGGCTTCTGCAGGCTTCAGCATCCACTTCACCATCCACTTCGGATCTCTC 804

Qy   2156 CTGAACTTCCCACGACGCTATCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGC 2215
               |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    805 CTGAACTCCCACGACGCTATCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGC 864

Qy   2216 CAAGCAAAATGCACACCCTTTTCTGTGGCAAATGTCTCTGTGCATGTGTGTACATATT 2275
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    865 CAAGCAAAATGCACACCCTTTTCTGTGGCAAATGTCTCTGTGCATGTGTGTACATATT 924

Qy   2276 AGAAAGGGAAGGAAGATGTTAAGATAATGTGGCCTGTGGTTACACAGGGTGCCTGCAGCGGTAA 2335
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    925 AGAAAGGGAAGGAAGATGTTAAGATAATGTGGCCTGTGGTTACACAGGGTGCCTGCAGCGGTAA 984

Qy   2336 TATATTTTAGAAATAATATCAAATAATCAACTAACTCCAATTTTAATCAATTATTA 2395
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    985 TATATTTTAGAAATAATATCAAATAATCAACTAACTCCAATTTTAATCAATTATTA 1044

Qy   2396 ATTTTTTTTCTTTTTAAAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTT 2455
               |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   1045 ATTTTTTTTCTTTTTAAAGAAAGCAGGCTTTTCTAGACTTTAAAGACTAAAGTCTTT 1104

Qy   2456 GGGAGGTCTCACGGTGTAGAGAGAGGAGCTTTGAGGCCACCCGGCCACCCGGCCACCCGGAG 2515
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1105 GGGAGGTCTCACGGTGTAGAGAGAGGAGCTTTGAGGCCACCCGGCCACCCGGAG 1164

Qy   2516 GAAATCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAAG 2575
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1165 GAAATCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAAG 1224

Qy   2576 GGATACCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTCAGTCAGCTCATCACAC 2635
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1225 GGATACCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTCAGTCAGCTCATCACAC 1284
```

```
Qy   2636  CCATTTCTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTC  2695
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1285  CCATTTCTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTC  1344

Qy   2696  TGTCTCTCCTGTTATCTCCCCCCCTTCTTCCCATCTCCATTCTT  2755
           ||||||||||||||  |||||||||||||||||||||||||||||
Db   1345  TGTCTCTCTCTG--TTATCTCCCCCCCTTCTTCCCATCTCCATTCTT  1403

Qy   2756  TTGAATTTCCTCATCCTCCATCTCAATCCGTATCTACGCA--CCCCCCCCCCCAGG  2813
           ||||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db   1404  TTGAATTTCCTCATCCTCCATCTCAATCCGTATCTACGCACCCCCCCCCCCCCCAGG  1463

Qy   2814  CAAAGCAGTGCTCTGAGTATCACATCACACAAAAGGAACAAAAGCGAAACACACAAAACCA  2873
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1464  CAAAGCAGTGCTCTGAGTATCACATCACACAAAAGGAACAAAAGCGAAACACACAAAACCA  1523

Qy   2874  GCCTCAACTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTT  2933
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1524  GCCTCAACTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGTGACTTGTCCTAGCGTT  1583

Qy   2934  TTGGAAGAGGAAAACAGGAACCCACCAACCAACCAATCAACCAAAGAAAAGAAAAAATTC  2993
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1584  TTGGAAGAGGAAAACAGGAACCCACCAACCAACCAATCAACCAAACAAAGAAAAGAAAAAATTC  1643

Qy   2994  CACAATGAAAGAAATGTATTTTGTCTTTTGGTGTATAAGCCATCAATATTCAG  3053
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1644  CACAATGAAAGAAATGTATTTTGTCTTTTGGTGTGTATAAGCCATCAATATTCAG  1703

Qy   3054  CAAAATGATTCCTTTCTTT-AAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGT  3112
           ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   1704  CAAAATGATTCCTTTCTTTCTTTAAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGT  1763

Qy   3113  TGTTGGCCCAGGGCGTTAAATTCACAGATTTTTTAACGAGAGAAAAACCACACAGAAGAGC  3172
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1764  TGTTGGCCCAGGGCGTTAAATTCACAGATTTTTTAACGAGAGAAAAACACACAGAAGAGC  1823

Qy   3173  TACCTCAGGTGTTTTTACCTCCAGCACCTTGCTCTGTGTTTCCCTTAGAGATTTTGTAAA  3232
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1824  TACCTCAGGTGTTTTTTACCTCCAGCACCTTGCTCTGTGTTTCCCTTAGAGATTTTGTAAA  1883

Qy   3233  GCTGATAGTTGGAGCATTTTTTATTTTTTTAATAAAATGAGTTGGAAAAAAATAAGA  3292
           ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   1884  GCTGATAGTTGGAGCATTTTTTTATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGA  1943

Qy   3293  TATCAACTGCCAGCCTGGAGCATTTTTTATTTCTTTTGGACAAACCTTGAAAATGTTTATTT  3352
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1944  TATCAACTGCCAGCCTGGAGCATTTTTTTATTTTTTATTTCTTTTGGACAAACCTTGAAAATGTTTATTT  2003

Qy   3353  CTTCCCTAGCCAAGAACCNATATGCCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT  3412
           ||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   2004  CTCCCGCTAGCCAAGAACCTATATGCCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT  2063
```

RESULT 8
ABX73494
ID   ABX73494 standard; DNA; 2290 BP.
XX
AC   ABX73494;
XX
DT   18-MAR-2003 (first entry)
XX
DE   Human novel polynucleotide #322.
XX
KW   Human; gene; ds; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW   haemostatic; antiarteriosclerotic.
XX
OS   Homo sapiens.
XX

CC   The invention relates to human novel polypeptides and their associated
CC   polynucleotides. The polypeptides and polynucleotides are useful in gene
CC   therapy for treating, inhibiting or preventing neural disorders, immune
CC   system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC   and multiple sclerosis), muscular disorders, respiratory diseaes (e.g.
CC   nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC   gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC   (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC   heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC   renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC   leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC   appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC   related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC   infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent
XX   human novel polynucleotides of the invention
SQ   Sequence 2290 BP; 707 A; 561 C; 491 G; 531 T; 0 U; 0 Other;

Query Match   58.3%;   Score 1989.8;   DB 7;   Length 2290;
Best Local Similarity   99.4%;
Matches 2028;   Conservative   0;   Mismatches   8;   Indels   4;   Gaps   3;
Pred. No. 0;

```
Qy   1376  CCCAGGCTGTGGGGCCATCATCGGGAAGAAGAGGGGCACACATCAAACAGCTGGCGAGAT  1435
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25    CCCAGCGTCCGGCGGCCATCATCGGGAAGAAGAGGGGCACACATCAAACAGCTGGCGAGAT  84

Qy   1436  TCGCCGGAGCCTCTATCAAGATTGCCCCTGCGAAGGCCCAGAGCGTCAGCAGGGATGG  1495
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   85    TCGCCGGAGCCTCTATCAAGATTGCCCCTGCGAAGGCCCAGAGCGTCAGCAGGGATGG  144

Qy   1496  TCATCATCACCGGGCCACCCGAAGCCCAGTTCAAGGCCCAGGACGGATCTTTGGGAAAC  1555
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   145   TCATCATCACCGGGCCACCCGAAGCCCAGTTCAAGGCCCAGGACGGATCTTTGGGAAAC  204

Qy   1556  TGAAAGAGAGAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGGAAGCGCATATCAGAG  1615
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   205   TGAAAGAGAGAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGGAAGCGCATATCAGAG  264

Qy   1616  TGCCCTTCCACAGCTGGCCGGGTGATTGGCAAGAGGTGGCAAGACCGTGAACGAACTGC  1675
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   265   TGCCCTTCCACAGCTGGCCGGGTGATTGGCAAGAGGTGGCAAGACCGTGAACGAACTGC  324

Qy   1676  AGAACTTAACCAGTGCAGAAGTCATCGGGCCACTTCTTTGCCTGACCCAAACGCCAGATGAAAATGAGG  1735
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   325   AGAACTTAACCAGTGCAGAAGTCATCGGGCCACTTCTTTGCCTGACCCGATGAAAATGAGG  384

Qy   1736  AAGTGATGATCGTCAGAATTATCGGGCACTTCTTTGCCTGACCCGATGAAGATCA  1795
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   385   AAGTGATGATCGTCAGAATTATCGGGCACTTCTTTGCCTGACCCGATGAAGATCA  444

Qy   1796  GGGAAATTGTACACACAGGTGAAGCAGGAGAGCAGCAGAGCAGAAATACCCTCAGGGAGTGCCTCAC  1855
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   445   GGGAAATTGTACACACAGGTGAAGCAGGAGAGCAGCAGAGCAGAAATACCCTCAGGGAGTCGCCTCAC  504

Qy   1856  AGCGGCAGCAAGTGAGGCTTCCACCAGAGGCACCAGCAAAACAAGCGGATGAATGTAGCCCCTTCC  1915
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   505   AGCGGCAGCAAGTGAGGCTTCCACCAGAGCGCACCAGCAAAACAAGCGGATGAATGTAGCCCCTTCC  564

Qy   1916  AACACCTGACAGAATGAGACCAAAACGCAGCGGAGCAAACCAAAGACCAT  1975
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   565   AACACCTGACAGAATGAGACCAAAACGCAGCGGAGCAAACCAAAGACCAT  624

Qy   1976  CTGAGGAATGAGAAGTCTGCGGAGGCGGCCAGGGACTTCTGCCGAGGCCCTGAGAACCCCA  2035
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   625   CTGAGGAATGAGAAGTCTGCGGAGGCGGCCAGGGACTTCTGCCGAGGCCCTGAGAACCCCA  684

Qy   2036  GGGGCCGAGGGAAGGGGGGGGAAGTCAGCCAGGTTGCCAGAACCACCGAGCCCCGCCTC  2095
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   685   GGGGGCCGAGGGAAGGGGGGGGGAAGTCAGCCAGGTTGCCAGAACCACCGAGCCCCGCCTC  744

Qy   2096  CCGCCCCCAGGGCTTCTGCAGGCTTCAGCCATCCACTTCACCATCCACTCGGATCTCTC  2155
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   745   CCGCCCCCAGGGCTTCTGCAGGCTTCAGCCATCCACTTCACCATCCACTCGGATCTCTC  804
```

```
Qy   2156  CTGAACTCCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGC  2215
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   805   CTGAACTCCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGC  864

Qy   2216  CAAGCAAAATGCACACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATATT  2275
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   865   CAAGCAAAATGCACACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATATT  924

Qy   2276  AGAAAGGAAGATGTTAAGATATGTGGGGTGGGGTACACAGGGTGCCTGCAGCCGGTAA  2335
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   925   AGAAAGGGAAGATGTTAAGATATGTGGCCTGTGGGGTACACAGGGTGCCTGCAGCCGGTAA  984

Qy   2336  TATATTTTAGAAATAATATATATCAAATAACTCAACTCCAATTTTAATCAATTATTA  2395
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   985   TATATTTTAGAAATAATATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTA  1044

Qy   2396  ATTTTTTTTTCTTTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTT  2455
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1045  ATTTTTTTTTCTTTTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGACTAAAGTCTTT  1104

Qy   2456  GGGAGGTCTCACGGTGTGAGAGGAGCGTTTGAGCACCCGACACAAATTCACCCAGGAGG  2515
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1105  GGGAGGTCTCACGGTGTGAGAGGAGCGTTTGAGCGCCACCCGACACAAAATTCACCCAGGAGG  1164

Qy   2516  GAAATCTTCGTCGGAAGGACACTCACGGCAGTTCTGATCACCTGTGTATGTCAACAGAAG  2575
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1165  GAAATCTTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTGTGTCAACAGAAG  1224

Qy   2576  GGATACCGTCTTCCTTGAAGAGGAAAACTTCTGTCACTCCTCATGCCTGCTAGCTCATCACAC  2635
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1225  GGATACCGTCTTCCTTGAAGAGGAAACTCTGTCACTCCTCATGCCTGCTCTAGCTCATCACAC  1284

Qy   2636  CCATTTCTTCTTTGCTTCACAGGTTTTTAAACTGGTTTTTTGCATACTGCTATATAAATTCTC  2695
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1285  CCATTTCTTCTTTGCTTCACAGGTTTTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTC  1344

Qy   2696  TGTCTCTCTCTGTTTATCTCTCCCCTTCTTCTTCTTCCATCTCATTCTTCTC  2755
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1345  TGTCTCTCTG-TTATCTCTCCCTTCTTCTTCTTCCATCTCATTCTTCTC  1403

Qy   2756  TTGAATTTCCTCATCCCTCGAGTATCAATCCGATCCAATCCGATGAAAATGAGG  2813
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1404  TTGAATTTCCTCATCCCTCGAGTATCAATCCGATCCAATCCGATGAAAATGAGG  1463

Qy   2814  CAAAGCAGTGCTCTGAGTATCACATCACACAAAGAGGAACAAAAAGGCGAAACACAACACAAACCA  2873
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1464  CAAAGCAGTGCTCTGAGTATCACATCACACAAAGAGGAACAAAAAGGCGAAACACAACACAAACCA  1523

Qy   2874  GCCTCAACTTACACTTGGTTACTCAAGAGCAGAGACAAGAGTCAATGTGCTACTTGTCCTAGCGTT  2933
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1524  GCCTCAACTTACACTTGGTTACTCAAGAGCAGAGACAAGAGTCAATGTGCTACTTGTCCTAGCGTT  1583

Qy   2934  TTGGAAGAGGAAAACAGGAACCCACCAAAACCACCAATCAACCAAACAAAGAAAAATTC  2993
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1584  TTGGAAGAGGAAAACAGGAACCCACCAACCACCAATCAACCAACCAAACAAAGAAAAATTC  1643

Qy   2994  CACAATGAAAGAAGAATGTATTTTGTCTTTTTGCATTTTGTCTTTTGCATTTTGGTGTATATTCAATATTCAG  3053
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1644  CACAATGAAAGAAGAATGTATTTTGTCTTTTTGCATTTTGGTGTATAAGCCATCAATATTCAG  1703

Qy   3054  CAAAATGATTCCTTTCTTT-AAAAAAAAAAAATGTGGAGGAAGTAGAAATTTACCAAGGT  3112
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1704  CAAAATGATTCCTTTCTTTTAAAAAAAAAAAATGTGGAGGAAGTAGAAATTTACCAAGGT  1763

Qy   3113  TGTTGGCCCAGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAGAGAAGC  3172
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1764  TGTTGGCCCAGGGCGTTAAATTCACAGATTTTTTTAACGAGAAAAACACACAGAGAGAAGC  1823

Qy   3173  TACCTCAGGTGTTTTTACCTCAGCCACCTTGCTCTGTGTTTCCCTTAGATTTTGTAAA  3232
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1824  TACCTCAGGTGTTTTTACCTCAGCCACCTTGCTCTGTGTTTCCCTTAGATTTTGTAAA  1883
```

```
Qy  3233  GCTGATAGTTGGAGCATTTTTTATTTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGA  3292
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1884  GCTGATAGTTGGAGCATTTTTTATTTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGA  1943

Qy  3293  TATCAACTGCCAGCTGGAGAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGT  3352
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1944  TATCAACTGCCAGCTGGAGAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGT  2003

Qy  3353  CTTCCGCTAGCCAAGAACCNATATGGCCTTCTTTTGGACAAAACCTTGAAAATGTTTATTT  3412
            ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db  2004  CTCCCGCTAGCCAAGAACCTATATGGCCTTCTTTTGGACAAAACCTTGAAAATGTTTATTT  2063
```

```
SQ  Sequence 2010 BP; 565 A; 552 C; 532 G; 361 T; 0 U; 0 Other;

Query Match          58.0%;  Score 1979.8;  DB 5;  Length 2010;
Best Local Similarity 99.4%;  Pred. No. 0;
Matches 1987;  Conservative 0;  Mismatches 12;  Indels 0;  Gaps 0;
```

```
Qy    35  GCCGGGGAGCCGCGGCGTCTCCGGCGAAGAGACGGATGATGAACAAGCTTACATCGGGA   94
            ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12  GCCGTAGGGAGCCGCGGCCTCAGGGAGGATGACGGATGATGAACAAGCTTACATCGGGA   71

Qy    95  ACCTGAGCCCCGCCGTCACCGCCGCCAGCTCTTTGGGACAGGAGAGCTGC  154
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    72  ACCTGAGCCCCGCCGTCACCGCCGCCAGCTCTTTGGGACAGGAGAGCTGC  131

Qy   155  CCCTGGCGCGGACAGGTCCTGCTGAAGTCCGGCTACGCCTTCGTGGACTACCCGACCAGA  214
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   132  CCCTGGCGCGGACAGGTCCTGCTGAAGTCCGGCTACGCCTTCGTGGACTACCCGACCAGA  191

Qy   215  ACTGGGCCATCCGCGCCATGCAGCACCCCTCGGGTAAAGTGGAATTGCATGGGAAAATCA  274
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   192  ACTGGGCCATCCGCGCCATGCAGCACCCCTCGGGTAAAGTGGAATTGCATGGGAAAATCA  251

Qy   275  TGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGGAGCAGGAAAAATTCAGATTCGAAACA  334
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   252  TGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGGAGCAGGAAAAATTCAGATTCGAAACA  311

Qy   335  TCCCTCCTCACCTGCAGTGGAGGTGTTGGATGGACTTTTGGCTCAATATGGACAGTGG  394
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   312  TCCCTCCTCACCTGCAGTGGAGGTGTTGGATGGACTTTTGGCTCAATATGGACAGTGG  371

Qy   395  AGAATGTGAACAAGTCAACACAGACACAGAAACCGCCGTTGTCAACGTCACATATGCAA  454
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   372  AGAATGTGAACAAGTCAACACAGACACAGAAACCGCCGTTGTCAACGTCACATATGCAA  431

Qy   455  CAAGAGAAGAAGCAAAAATAGCCATGGGAGAAGAGGAAGCTAAGCGGCATCAGTTTGAGAACTACT  514
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   432  CAAGAGAAGAAGCAAAAATAGCCATGGGAGAAGAGGAAGCTAAGCGGCATCAGTTTGAGAACTACT  491

Qy   515  CCTTCAAGATTCCTACATCCCGATGAAGAGGTGAGCTGAGCTCCCCTTCGCCCCCCAGCGAG  574
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   492  CCTTCAAGATTCCTACATCCCGATGAAGAGGTGAGCTGAGCTCCCCTTCGCCCCCCAGCGAG  551

Qy   575  CCCAGCGTGGGGACCACTCTTCCCGGGAGCAAGGCCACGCCCAGCCCGCCCACTTCTCAGG  634
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   552  CCCAGCGTGGGGACCACTCTTCCCGGGAGCAAGGCCACGCCCAGCCCGCCCACTTCTCAGG  611

Qy   635  CCAGACAGATTGATTTCCCGCTCGGATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCA  694
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   612  CCAGACAGATTGATTTCCCGCTCGGATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCA  671

Qy   695  TCGGAAAGAGGGCTTGACCATAAGAGAACATCACTCAAGCAGACCCAGTCCGGGTAGATA  754
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   672  TCGGAAAGAGGGCTTGACCATAAGAGAACATCACTCAAGCAGACCCAGTCCGGGTAGATA  731

Qy   755  TCCATAGAAAAGAAGAACTTCTGAGCTGCAGAGAAGCTGTCACCATCCATGCCACCCCAG  814
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   732  TCCATAGAAAAGAAGAACTTCTGAGCTGCAGAGAAGCCTGTCACCATCCATGCCACCCCAG  791

Qy   815  AGGGGACTTCTGAGCATGCGCATGATGATTCTTGAAATCATGCAGAAGAGGCAGATGAGA  874
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   792  AGGGGACTTCTGAGCATGCGCATGATGATTCTTGAAATCATGCAGAAGAGGCAGATGAGA  851

Qy   875  CCAAACTAGCCGAAGAGATTCCTCGAAAATCTTGGCACACAATGGCTTGGTTGGAAGAC  934
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   852  CCAAACTAGCCGAAGAGATTCCTCGAAAATCTTGGCACACAATGGCTTGGTTGGAAGAC  911

Qy   935  TGATTGGAAAGAGAGGCAGAAATTGAAGAAATTGAACATGAAACAGGGACCAAGATAA  994
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   912  TGATTGGAAAGAGAGGCAGAAATTGAAGAAATTGAACATGAAACAGGGACCAAGATAA  971

Qy   995  CAATCTCATCTTTGCAGGATTTGAGCATATACACCCGAAAAGAACCATCACTGTGAAGG  1054
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   972  CAATCTCATCTTTGCAGGATTTGAGCATATACACCCGAAAAGAACCATCACTGTGAAGG  1031
```

```
RESULT 9
AAS70981
ID  AAS70981 standard; cDNA; 2010 BP.
XX
AC  AAS70981;
XX
DT  13-FEB-2002 (first entry)
XX
DE  DNA encoding novel human diagnostic protein #6785.
XX
KW  Human; chromosome mapping; gene therapy; forensic;
KW  food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS  Homo sapiens.
XX
PN  WO200175067-A2.
XX
PD  11-OCT-2001.
XX
PF  30-MAR-2001; 2001WO-US008631.
XX
PR  31-MAR-2000; 2000US-00540217.
PR  23-AUG-2000; 2000US-00649167.
XX
PA  (HYSB-) HYSEQ INC.
XX
PI  Drmanac RT, Liu C, Tang YT;
XX
DR  WPI; 2001-639362/73.
DR  P-PSDB; ABG06794.
XX
PT  New isolated polynucleotide and encoded polypeptides, useful in
PT  diagnostics, forensics, gene mapping, identification of mutations
PT  responsible for genetic disorders or other traits and to assess
PT  biodiversity.
XX
PS  Claim 1; SEQ ID NO 6785; 103pp; English.
XX
CC  The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC  sequences.(I) is useful as hybridisation probes, polymerase chain
CC  reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC  and in recombinant production of (II). The polynucleotides are also used
CC  in diagnostics as expressed sequence tags for identifying expressed
CC  genes. (I) is useful in gene therapy techniques to restore normal
CC  activity of (II) or to treat disease states involving (II). (II) is
CC  useful for generating antibodies against it, detecting or quantitating a
CC  polypeptide in tissue, as molecular weight markers and as a food
CC  supplement. (II) and its binding partners are useful in medical imaging
CC  of sites expressing (II). (I) and (II) are useful for treating disorders
CC  involving aberrant protein expression or biological activity. The
CC  polypeptide and polynucleotide sequences have applications in
CC  diagnostics, forensics, gene mapping, identification of mutations
CC  responsible for genetic disorders or other traits to assess biodiversity
CC  and to produce other types of data and products dependent on DNA and
CC  amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC  coding sequences of the invention. Note: The sequence data for this
CC  patent did not appear in the printed specification, but was obtained in
CC  electronic format directly from WIPO at
CC  ftp.wipo.int/pub/published_pct_sequences
XX
```

```
XX
AC   ACA90176;
XX
DT   10-JUL-2003  (first entry)
XX
DE   cDNA encoding novel human protein NOV14a.
XX
KW   Human; Cytostatic; DAPK3-Agonist; DAPK3-Antagonist; cancer; NOV; gene;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO200301571-A2.
XX
PD   17-APR-2003.
XX
PF   02-OCT-2002; 2002WO-US031357.
XX
PR   05-OCT-2001; 2001US-0327454P.
PR   09-OCT-2001; 2001US-0327917P.
PR   09-OCT-2001; 2001US-0328029P.
PR   09-OCT-2001; 2001US-0328056P.
PR   12-OCT-2001; 2001US-0328849P.
PR   15-OCT-2001; 2001US-0329414P.
PR   17-OCT-2001; 2001US-0330142P.
PR   22-OCT-2001; 2001US-0341058P.
PR   24-OCT-2001; 2001US-0343629P.
PR   29-OCT-2001; 2001US-0349575P.
PR   01-NOV-2001; 2001US-0346357P.
PR   25-JUN-2002; 2002US-0391342P.
PR   01-OCT-2002; 2002US-0262445.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Alsobrook JP, Burgess CE, Catterton E, Chant JS, Chaudhuri A;
PI   Edinger SR, Gerlach VL, Giot L, Gorman L, Guo X, Kekuda R;
PI   Mezes PS, Millet I, Ooi CE, Patturajan M, Rieger DK, Spytek KA;
PI   Taupier RJ, Zerhusen BD, Zhong H, Zhong M;
XX
DR   WPI; 2003-381704/36.
DR   P-PSDB; ABU89799.
XX
PT   New DAPK3 polypeptide, useful for preparing a composition for treating or
PT   preventing e.g., cancer.
XX
PS   Claim 20; Page 129; 253pp; English.
XX
CC   The invention describes an isolated polypeptide comprising any of 33 90-
CC   1273 amino acid sequences (I) given in the specification or its mature
CC   form, a sequence that is at least 95 % identical to (I), or a sequence
CC   comprising one or more conservative substitutions in the amino acid
CC   sequence of (I). The polypeptide is useful for preparing a composition
CC   for treating or preventing e.g. cancer. This sequence encodes a novel
CC   human NOV protein
XX
SQ   Sequence 1707 BP; 492 A; 443 C; 442 G; 330 T; 0 U; 0 Other;
```

Query Match   41.7%; Score 1423.8; DB 7; Length 1707;
Best Local Similarity   88.4%;
Matches 1625; Conservative 0; Mismatches 82; Indels 132; Gaps 2;

```
Qy     66 ACGGATGATGAACAAGTTTACATCGGGAACTGAGCCCCGTCACCGCGACGACCT 125
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ACGGATGATGAACAAGTTTTCATCGGGAACTGAGCCCCGTCACCGCGAGACCT 60

Qy    126 CCGGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGGGGGGACAGGGTCCTGCTGCTCAAGTCCGG 185
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCGGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGGGGGGACAGGGTCCTGCTGCTCAAGTCCCG 120

Qy    186 CTTACGCCTTCGTGACTACCCCGACCAGAACTGGGCCATCCGGCGACGACTGGGCCACCATCCGACGACCCTCTC 245
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTTACGCCTTCGTGACTACCCCGACCAGAACTGGGCCATCCGGCCACCATCCGACGACCCTCTC 180
```

```
Qy   1055 GCACAGTTGAGGCCTGTGCCAGTGCTGAGATAGAGATTATGAAGAAGCTGCGTGAGGCCT 1114
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1032 GCACAGTTGAGGCCTGTGCCAGTGCTGAGATAGAGATTATGAAGAAGCTGCGTGAGGCCT 1091

Qy   1115 TTGAAAATGATATGCTGGCTGTTAACCAACAAGCCAATCTGATCCCAGGGTTGAACCTCA 1174
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1092 TTGAAAATGATATGCTGGCTGTTAACCAACAAGCCAATCTGATCCCAGGGTTGAACCTCA 1151

Qy   1175 GCGCACTTGGCATCTTTCAACAGGACTGTCCGTGCTATCTCCACCAGCAGGGCCCCGCG 1234
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1152 GCGCACTTGGCATCTTTTCAACAGGACTGTCCGTGCTATCTCCACCAGCAGGGCCCCGCG 1211

Qy   1235 GAGCTCCCCCGCTGCCCCCTACCACCCCTTCACTACCCACTCCGGATACTTCTCCAGCC 1294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1212 GAGCTCCCCCGCTGCCCCCTACCACCCCTTCACTACCCACTCCGGATACTTCTCCAGCC 1271

Qy   1295 TGTACCCCCATCACCAGTTTGGCCCGTTCCCGTTCCGCATCATCATCACTCTTATCCAGAGCAGGAGA 1354
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1272 TGTACCCCCATCACCAGTTTGGCCCGTTCCCGTTCCGCATCATCATCACTCTTATCCAGAGCAGGAGA 1331

Qy   1355 TTGTGAATCTCTTCATCCCAACCCAGGCTGTGGGCGCCATCATCGGGAAGAAGGGGGCAC 1414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1332 TTGTGAATCTCTTCATCCCAACCCAGGCTGTGGGCGCCATCATCGGGAAGAAGGGGGCAC 1391

Qy   1415 ACATCAAACAGCTGGCGAGATTCGCCGGAGCCCTCTATCAAGATTGCCCCTGCCGGAAGGCC 1474
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1392 ACATCAAACAGCTCGGCGAGATTCGCCGGAGCCCTCTATCAAGATTGCCCCTGCCGGAAGGCC 1451

Qy   1475 CAGAGCTCAGCGAAAGGATGTGTCATCATCACCGGCCACCGGAGCCCAGTTCAAGGCCC 1534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1452 CAGAGTCAGCGAAAGGATGTGTCATCATCACCGGCCACCGGAGCCCAGTTCAAGGCCC 1511

Qy   1535 AGGGACGGAGATCTTTGGGGAAACTGAAAGAGGGAAAACTTCTTTAACCCAAAGAGAAGAAGTGA 1594
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1512 AGGGACGACGGGATCTTTGGGAAACTGAAAGAGGGAAAACTTCTTTAACCCAAAGAGAAGAAGTGA 1571

Qy   1595 AGCTGGAAGCCATATCAGAGTGCCCTCTTCCACAGTGCCTGGCCGGTGATTGGCAAAGTG 1654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1572 AGCTGGAAGCCATATCAGAGTGCCCTCTTCCACAGTGCCTGGCCGGTGATTGGCAAAGTG 1631

Qy   1655 GCAAGACCGTGAACGAACTGCAGAACTTAACCAGTGCAGAAGTCATCGTCGCCTCGTGACC 1714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1632 GCAAGACCGTGAACGAACTGCAGAACTTAACCAGTGCAGAAGTCATCGTCGCCTCGTGACC 1691

Qy   1715 AAACGCCAGATGAAAATGAGGAAGTGATCGTCAGAATTATCGGGCACTCTTTGCTAGCC 1774
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1692 AAACGCCAGATGAAAATGAGGAAGTGATCGTCAGAATTATCGGGCACTCTTTGCTAGCC 1751

Qy   1775 AGACTGCACAGCGCAAGATCAGGGAAATTGTACAACAGTGAGCAGCAGGAGCAGAAAT 1834
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1752 AGACTGCACAGCGCAAGATCAGGGAAATTGTACAACAGTGAGCAGCAGGAGCAGAAAT 1811

Qy   1835 ACCCTCAGGGAGTCGCCTCACAGCGCGCAGCAAGTGAGGCTCCCACAGCCACCAGCAAAACA 1894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1812 ACCCTCAGGGAGTCGCCTCACAGCGCGCAGCAAGTGAGGCTCCCACAGCCACCAGCAAAACA 1871

Qy   1895 ACGGATGAATGATAGCCCTGACACACTGACAGAATGAGACCAAACGCAGCCAGCCAGAT 1954
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1872 ACGGATGAATGATAGCCCTTCCAACACCTGACACGAATGAGACCAAACGCAGCCAGCCAGAT 1931

Qy   1955 CGGGAGCAAACCAAAGACCATCTGAGGAATGAGAGTCTGCGGAGGCGCCCAGGGACTCT 2014
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1932 CGGGAGCAAACCAAAGACCATCTGAGGAATGAGAGTCTTTCATCGGGAGGCGCCCAGGGACTCT 1991

Qy   2015 GCCGAGGCCCTGAGAACCC 2033
          |||||||||||||||||||
Db   1992 GCCGAGGCCCTGAGAACCC 2010
```

RESULT 10
ACA90176
ID   ACA90176 standard; cDNA; 1707 BP.

```
Qy    246  GGGTAAAGTCGAAGTGCATGGGAAAATCATGGGAAGTTGATTACTCAGTCTCTAAAAAGCT  305
                |||| |||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db    181  GGGTCAAGTCGAAATGCATGGGAAAATCATGGGAAGTTGATTATTCAGTCTCTATAAAGCT  240

Qy    306  AAGGAGGACGAGAAATTCAGATTGAAACATCCCTCCTCCACCTGCAGTTGGGAGGTGTTGGA  365
                |||||| |||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    241  AAGGAGCAGAACATTCCGATTCCGATTCGAAATATCCCTCCTCCACCTGCAGTGGGAGGTGTTGGA  300

Qy    366  TGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGGAACAGAGTCAACACAGACACAGA  425
                ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db    301  TGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGGAACAGAGTCAACACAGACACAGA  360

Qy    426  AACCCCGTTGTCAAGTCACATATGCAACAAGAGAAGAGAAAATAGCCATGGAGAA       485
                |||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    361  GACCCCGTTGTCAAGTCACATATGCAACAAGAGAAGAGAAAAATAGCCATGAAGAA       420

Qy    486  GCTAAGGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTCTACATCCCGGATGAGAA    545
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  GCTAAGGCGGGCATCAGTTTGAGAACCACTACTTCAAGATTTCAAGATTCCTACATCCCGGATGACGA  480

Qy    546  GGTGAGCTCCCCTTCGCCCCAGCGTGGGGACCCACTCTTCCCGGGACCA          605
                ||||||||||||||||||||||||||||||||||||||||||||| |||
Db    481  GGTGAGCTCCCCTTCGCCCCAGCGTGGGGACCCACTCTTCCTTGGGACCA          540

Qy    606  AGGGCCACGCCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTCCCGTCGCGGATCCT  665
                ||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db    541  AGGGCCAAGCCCCTGGGGGCTCTTCTCAGGCCAGACAGATTGATTCCCACTCGCGTGTCCT  600

Qy    666  GGTCCCACCCAGTTTGTTGTGCCATCATCGGGGCTTGACCATAAAGAACAT       725
                |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  GTTCCCACCCAGTTTGTTGTGCCATCATCGGGGCTTGACCATAAAGAACAT       660

Qy    726  CACTAAGCAGACCCAGTCCCGGGTAGTATCCATAGAGAAAAGAGAAACTTCTGAGCTGCGAGA  785
                ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db    661  CACTAAGCAGACGAGCCCGGGTAGACATCTATAGACAAGAGAAACTCCAGAGCTGCGAGA  720

Qy    786  GAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGATGATTCT   845
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  GAAGCCTGTCACCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATGATGATTCT   780

Qy    846  TGAAATCATGCAGAAGAGAGGCAGATGAGACCCAAACTAGCCGAAGAGATTCCTCTGAAAAT  905
                |||| |||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db    781  TGAAATAATGCAGAAGAGAGGCAGATGAGACCCAAACTAGCCGAAGAGATTCCTCTGAAAAT  840

Qy    906  CTTGGCACACAATGGCTTGGTTGGTTGGAAGACTGATTGGAAGAGAAGGCAGAAATTTGAAGAA  965
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  CTTGGCCCACAATGGCTTGGTTGGTTGGAAGACTGATTGGAAGAGAAGGCAGAAATTTGAAGAA  900

Qy    966  AATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATATA  1025
                ||||||| |||||||||||||||| ||| |||||||||| |||||||||||||||||||
Db    901  AAATGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTCGCGAGATTTGAGCATATA  960

Qy   1026  CAACCCGGAAAGAACAACCATCACTGTGAAGGGCACAGTTGGCAGGCCTGTCCCAGTGCTGAGAT  1085
                ||||||||||||||||||||||||||||||||||||||| ||| ||||||||||||||||
Db    961  CAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGGCGAGTCTGTGCCAGTGCTGAGAT  1020

Qy   1086  AGAGATTATGAAGAAGCTGCCTGAGGCCTTTGAAAATGATGCTGGCTGTGTAACCAACA  1145
                |||||||||||||||||||||||||||||||||||||| ||| |||||
Db   1021  AGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATACGCTGACTGTTTA------  1074

Qy   1146  AGCCAATCTGATCCCCAGGGTTGAACCTTCAGCGCACTTGGCATCTTTTCAACAGGACTGTC  1205
                --------------------------------------------------------------
Db   1075  ------------------------------------------------------------  1074

Qy   1206  CGTGCTATCTCCACCACCAGCAGGCCCCGCGGAGCTCCCCGCCCTGCCCCTACCCCCTT  1265
                ------------------------------------------------------------
Db   1075  ------------------------------------------------------------  1074

Qy   1266  CACTACCCCACTCCGGATACTTCTCCAGCCTCCGGATACCCGTACCCCCATCACCAGTTTTGGCCCGTTCCC  1325
                |||||||||||            ||||||||||||||||||||||||||||||||||
Db   1075  ---TACCCACTTCGGATACTTCTCCAGCCTCCGGATACCCGTACCCCCATCGCCAGTTTGGCCCGTTCCC  1131

Qy   1326  GCATCATCACTCTTATCCAGAGCAGCAGGAGAGATTGTGAATCTCTTCATCCCAACCCCAGGCTGT  1385
```

```
Db   1132  GCATCATCACTCTTATCCAGAGCAGCAGGAGAGATTGTTCAATCTCTTCATCCCAACCCCAGGGGTGT  1191

Qy         GGGCGCCATCATCGGGAAGAAGGGGCACACTCAAACAGTCGGCGAGATTGCCGAGAGC  1445
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1386  GGGCGCCATCATCGGGAAGAAGGGGCACACTCAAACAGTCGGCGAGATTGCGTGGGAGC  1251

Qy   1446  CCTATCAAGATGCCCCTGCCGGAAGGCCCAGACGTCAGCGAACGTCATCACAC  1505
                 ||||||||||||||||||||||||||||||||           |||||||||||||||
Db   1252  CTCCATCAAGATCGCCCCTGCCGAGATCCCC---CCTGCGTCAGCGGAAGGTCATCATCAC  1308

Qy   1506  CGGGCCACCGGAAGCCCCAGTTCAAGGCCCCAGGGACGACGGATCTTTGGGAAACTGAAAGAGGA  1565
                ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   1309  CTGGCCACCGGAATCCCAGTTCAAGGCCCCAGGGACGACGGATCTTTGGGAAACTGAAAAGAAGA  1368

Qy   1566  AAACTTCTTAACCCCAAAGAAGAAGTGAAGCTGGAAGCGCATATCAGAGTGCCCCTTC  1625
                ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1369  AAACTTTTTTAACCCCAAAGAAGACGTGAAGCGTGAAACCCATATCAGAGTGCCCCTCTTC  1428

Qy   1626  CACAGCTGCCGGGTGATTGGCAAAGGTGGCAAGACGTGACAACGAACTGCAGAACTAAAC  1685
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1429  CACCGCTGCCGGGTGATTGGCAAAGGTGGCAAGACCGTGAATGAACTGCAGAATTTAAT  1488

Qy   1686  CAGTGCAGAAGTCATCGTGCCTCGTGACAAAGCGCCAGATGAAAATGAGGAGAGTGATCGT  1745
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1489  CAGTGCAGAAGTCATCGTGCCCTCGTGACCAAATGCGCCAAGATCAGGAATGATCGT  1548

Qy   1746  CAGAATTATCGGGCACTCTTTTGCTAGCCAGACTGCACAGCCAAGATCAGGGAAAATTGT  1805
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1549  CAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCCAAGATCAGGGAAAATTGT  1608

Qy   1806  ACAACAGGTGAAGCAGCAGGAGAGCAGAGAAATACCCTCAGGGAGTGCGCTCCACAGCGCAGCAA  1865
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1609  ACAACAGGTGAAGCAGCAGGAGAGCAGAGAAATACCCTCAGGGAGTGCGCTCCACAGCGCAGCAA  1668

Qy   1866  GTGAGGCTCCACAGGCACCAGCAAACAACGGATGAAT  1904
                |||||||||||||||||||||||||||||||||||||
Db   1669  GTGAGGATCCCACAGGCACAAGCAAACAACGGAAGAAT  1707
```

RESULT 11
AAS26570

PR   17-MAR-2000;   2000US-0190076P.
PR   18-APR-2000;   2000US-0198123P.
PR   19-MAY-2000;   2000US-0205515P.
PR   07-JUN-2000;   2000US-0209467P.
PR   28-JUN-2000;   2000US-0214886P.
PR   30-JUN-2000;   2000US-0215135P.
PR   07-JUL-2000;   2000US-0216647P.
PR   07-JUL-2000;   2000US-0216880P.
PR   11-JUL-2000;   2000US-0217487P.
PR   11-JUL-2000;   2000US-0217496P.
PR   14-JUL-2000;   2000US-0218290P.
PR   26-JUL-2000;   2000US-0209963P.
PR   26-JUL-2000;   2000US-0209964P.
PR   14-AUG-2000;   2000US-0224518P.
PR   14-AUG-2000;   2000US-0224519P.
PR   14-AUG-2000;   2000US-0225213P.
PR   14-AUG-2000;   2000US-0225214P.
PR   14-AUG-2000;   2000US-0225266P.
PR   14-AUG-2000;   2000US-0225267P.
PR   14-AUG-2000;   2000US-0225268P.
PR   14-AUG-2000;   2000US-0225270P.
PR   14-AUG-2000;   2000US-0225447P.
PR   14-AUG-2000;   2000US-0225557P.
PR   14-AUG-2000;   2000US-0225758P.
PR   14-AUG-2000;   2000US-0225759P.
PR   18-AUG-2000;   2000US-0226279P.
PR   22-AUG-2000;   2000US-0226681P.
PR   22-AUG-2000;   2000US-0226868P.
PR   22-AUG-2000;   2000US-0227182P.
PR   23-AUG-2000;   2000US-0227009P.
PR   30-AUG-2000;   2000US-0228924P.
PR   01-SEP-2000;   2000US-0229287P.
PR   01-SEP-2000;   2000US-0229343P.
PR   01-SEP-2000;   2000US-0229344P.
PR   01-SEP-2000;   2000US-0229345P.
PR   05-SEP-2000;   2000US-0229509P.
PR   05-SEP-2000;   2000US-0229513P.
PR   06-SEP-2000;   2000US-0230437P.
PR   06-SEP-2000;   2000US-0230438P.
PR   08-SEP-2000;   2000US-0231242P.
PR   08-SEP-2000;   2000US-0231243P.
PR   08-SEP-2000;   2000US-0231244P.
PR   08-SEP-2000;   2000US-0231413P.
PR   08-SEP-2000;   2000US-0231414P.
PR   08-SEP-2000;   2000US-0232080P.
PR   08-SEP-2000;   2000US-0232081P.
PR   12-SEP-2000;   2000US-0231968P.
PR   14-SEP-2000;   2000US-0232397P.
PR   14-SEP-2000;   2000US-0232398P.
PR   14-SEP-2000;   2000US-0232399P.
PR   14-SEP-2000;   2000US-0232400P.
PR   14-SEP-2000;   2000US-0232401P.
PR   14-SEP-2000;   2000US-0233063P.
PR   14-SEP-2000;   2000US-0233064P.
PR   14-SEP-2000;   2000US-0233065P.
PR   21-SEP-2000;   2000US-0234223P.
PR   21-SEP-2000;   2000US-0234274P.
PR   25-SEP-2000;   2000US-0234997P.
PR   26-SEP-2000;   2000US-0234998P.
PR   27-SEP-2000;   2000US-0235484P.
PR   27-SEP-2000;   2000US-0235834P.
PR   29-SEP-2000;   2000US-0235836P.
PR   29-SEP-2000;   2000US-0236327P.
PR   29-SEP-2000;   2000US-0236367P.
PR   29-SEP-2000;   2000US-0236368P.
PR   29-SEP-2000;   2000US-0236369P.
PR   29-SEP-2000;   2000US-0236370P.
PR   02-OCT-2000;   2000US-0236802P.
PR   02-OCT-2000;   2000US-0237037P.
PR   02-OCT-2000;   2000US-0237038P.
PR   02-OCT-2000;   2000US-0237039P.
PR   02-OCT-2000;   2000US-0237040P.
PR   13-OCT-2000;   2000US-0239935P.

PR   13-OCT-2000;   2000US-0239937P.
PR   20-OCT-2000;   2000US-0240960P.
PR   20-OCT-2000;   2000US-0241221P.
PR   20-OCT-2000;   2000US-0241785P.
PR   20-OCT-2000;   2000US-0241786P.
PR   20-OCT-2000;   2000US-0241787P.
PR   20-OCT-2000;   2000US-0241808P.
PR   20-OCT-2000;   2000US-0241809P.
PR   20-OCT-2000;   2000US-0241826P.
PR   01-NOV-2000;   2000US-0244617P.
PR   08-NOV-2000;   2000US-0244744P.
PR   08-NOV-2000;   2000US-0246475P.
PR   08-NOV-2000;   2000US-0246476P.
PR   08-NOV-2000;   2000US-0246477P.
PR   08-NOV-2000;   2000US-0246478P.
PR   08-NOV-2000;   2000US-0246523P.
PR   08-NOV-2000;   2000US-0246524P.
PR   09-NOV-2000;   2000US-0246525P.
PR   08-NOV-2000;   2000US-0246526P.
PR   08-NOV-2000;   2000US-0246527P.
PR   08-NOV-2000;   2000US-0246528P.
PR   08-NOV-2000;   2000US-0246532P.
PR   08-NOV-2000;   2000US-0246609P.
PR   08-NOV-2000;   2000US-0246610P.
PR   08-NOV-2000;   2000US-0246611P.
PR   08-NOV-2000;   2000US-0246613P.
PR   17-NOV-2000;   2000US-0249207P.
PR   17-NOV-2000;   2000US-0249208P.
PR   17-NOV-2000;   2000US-0249209P.
PR   17-NOV-2000;   2000US-0249210P.
PR   17-NOV-2000;   2000US-0249211P.
PR   17-NOV-2000;   2000US-0249212P.
PR   17-NOV-2000;   2000US-0249213P.
PR   17-NOV-2000;   2000US-0249214P.
PR   17-NOV-2000;   2000US-0249215P.
PR   17-NOV-2000;   2000US-0249216P.
PR   17-NOV-2000;   2000US-0249217P.
PR   17-NOV-2000;   2000US-0249218P.
PR   17-NOV-2000;   2000US-0249244P.
PR   17-NOV-2000;   2000US-0249245P.
PR   17-NOV-2000;   2000US-0249264P.
PR   17-NOV-2000;   2000US-0249265P.
PR   17-NOV-2000;   2000US-0249297P.
PR   17-NOV-2000;   2000US-0249299P.
PR   01-DEC-2000;   2000US-0249300P.
PR   01-DEC-2000;   2000US-0250160P.
PR   05-DEC-2000;   2000US-0250391P.
PR   05-DEC-2000;   2000US-0251030P.
PR   05-DEC-2000;   2000US-0251988P.
PR   06-DEC-2000;   2000US-0256719P.
PR   08-DEC-2000;   2000US-0251479P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
PR   08-DEC-2000;   2000US-0251989P.
PR   11-DEC-2000;   2000US-0251990P.
PR   05-JAN-2001;   2000US-0254097P.
PR   05-JAN-2001;   2001US-0256678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-488783/53.
DR   P-PSDB; AAU16583.
XX
PT   New nucleic acid molecules encoding 461 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives.
XX
PS   Claim 1; SEQ ID NO 749; 980pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their

```
CC  encoded secreted proteins. The nucleic acids and proteins are used to
CC  prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC  rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC  in diagnosing a pathological condition or susceptibility to a
CC  pathological condition. Antibodies to the proteins can also be used in
CC  alleviating symptoms associated with the proteins and in diagnostic
CC  immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC  (ELISA). Disorders which are diagnosed or treated include autoimmune
CC  diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC  neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC  arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC  nervous system disorders e.g. Alzheimer's disease, infections caused by
CC  bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC  and many other disorders lised in the specification. The polypeptides
CC  can also be used to aid wound healing and epithelial cell proliferation,
CC  to prevent skin aging due to sunburn, to maintain organs before
CC  transplantation, for supporting cell culture of primary tissues, to
CC  regenerate tissues and in chemotaxis. The polypeptides can also be used
CC  as a food additive or preservative to increase or decrease storage
CC  capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC  minerals, cofactors and other nutritional components. The present
CC  sequence encodes a novel secreted protein of the invention. Note: The
CC  sequence data for this patent did not form part of the printed


Query Match             32.1%;  Score 1096.8;  DB 4;  Length 1186;
Best Local Similarity   98.4%;  Pred. No. 1.3e-280;
Matches 1145;  Conservative  3;  Mismatches  12;  Indels  4;  Gaps  4;

Qy   1376 CCCAGGCTGTGGGGCGCCATCATCGGGAAGAAGGGGGCCACACATCAAAAACAGCTGCGAGAT 1435
               | | |||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db     25 CCCACGCGTCCGGCGCCATCATCGGGAAGAAGGGGGCCACACATCAAAAACAGCTGCGAGAT   84

Qy   1436 TCGCCGGAGCCTCTATCAAGATTGCCCTCGGAAGGCCCAGAACGTCAGCGAAAGGATGG 1495
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     85 TCGCCGGAGCCTCTATCAAGATTGCCCTCGGAAGGCCCAGAACGTCAGCGAAAGGATGG  144

Qy   1496 TCATCATCACCGGCACCGGAAGCCGAGCCAGTTCAAGGCTCAAGGACGGATCTTTGGGAAAC 1555
          |||||||||||||||  || :::||||| ||||||||| |||||||||||||||||||||
Db    145 TCATCATCACCGGCGCCA-CGSAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAAC  203

Qy   1556 TGAAAGAGGAAAACTTCTTTAACCCCAAAGAAGAGTGAAGCTGAAGGCCATATCAGAG 1615
          |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    204 TGAAAGAGGAAAACTTCTTTAACCCCAAAGAAGAGTGAAGCTGAACCCATTATCAGAG  263

Qy   1616 TGCCCTCTTCCACAGCTGGCCGGGTGATTGCAAGGTGGCAAAGCCGTGAACGAACTGC 1675
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    264 TGCCCTCTTCCACAGCTGGCCGGGTGATTGCAAGGTGGCAAAGCCGTGAACGAACTGC  323

Qy   1676 AGAACTTAACCAGTGCAGAAGTCGCAGAATGAGAGCTCATCGTGCCTCGTGACCAAAGCGCCAGATGAAAATGAGG 1735
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    324 AGAACTTAACCAGTGCAGAAGTCGCAGAATGAGAGCTCATCGTGCCTCGTGACCAAAGCGCCAGATGAAAATGAGG  383

Qy   1736 AAGTGATCGTCAGAATTATCGGGACACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCA 1795
          ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    384 AAGTGATCGTCAGAATTATCGGG-CACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCA  443

Qy   1796 GGGAAATTG-TACAACAGGTGAAGCAGCAGCAGAAATACCCTCAGGGAGTCGCCTCA 1854
          |||||||| ||||||||||||||||||||||  |||||||||||||||||||||||
Db    444 GGGAAATTGTTACAACAGGTGAAGCAGCAGCAGCAGAAATACCCTCAGGGAGTCGCCTCA  503

Qy   1855 CAGCCGCAGCAAGTGAGGCTCCCACAGGCACCAGCAAAACAACGATGAATGTAGCCCTTC 1914
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    504 CAGCCGCA-NAAGTGAGGCTCCCACAGGCACCAGCAAAACAACGATGAATGTAGCCCTTC  562

Qy   1915 CAACACCTGACAGAATGAGACAACAGCCAAACCAGCCAGATCGGAGGAGCAAACCAAAGACCA 1974
          |||||||||||| |||||||||||||||||||||||||||||| ||||||||||||||||||
Db    563 CAACACCTGACACGAATGAGACAACAGCCAAACCAGCCAGATCGGGAGGAGCAAACCAAAGACCA  622

Qy   1975 TCTGAGGAATGAGAAGTTCTGCCGAGGACTCTGCCGAGGCCCCTGAGACTTTGCCGAGAACCCC 2034
          ||||||||||||||||||||||||||||||||||||||||||||||             ||||
Db    623 TCTGAGGAATGAGAAGTTCTGCCGAGGACTCTGCCGAGGCCCCTGAGAACCCC  682

Qy   2035 AGGGGCCGAGGAGGGGGGGCGGGGAAGGTCAGCCAGGTTTGCCAGAACCACCGAGCCCGCCT 2094
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    683 AGGGGCCGAGGAGGGGGGGCGGGGAAGGTCAGCCAGGTTTGCCAGAACCACCGAGCCCGCCT  742

Qy   2095 CCCGCCCCCAGGGCTTCTGCGACGGCTTCAGCCAT-CCACTTCACCATCCACTCGGATCTC 2153
          |||||||||||||||| ||||||||||||||||| |||||||||||||||||||||||||
Db    743 CCCGCCCCCAGGGCTTCGCGACGGCTTCAGCNATNCCACTTCACCATCCACTCGGATCTC  802

Qy   2154 TCCTGAACTCCACGAGCTATCCCTTTAGTTGAACTAACATAGGTGAACCTGTTCAAA 2213
          |||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    803 TCCTGAACTCCACGAGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAA  862

Qy   2214 GCCAAGCAAATGCACACCCTTTTCTGTGGCAAATCGTCTGTACATGTGTACATA 2273
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    863 GCCAAGCAAATGCACACCCTTTTCTGTGGCAAATCGTCTGTACATGTGTACATA  922

Qy   2274 TTAGAAAGGGAAGATGTTAAGATATGTGGCCTGTGGGTACACAGGGTCCTGCAGCGGT 2333
          |||||||||||||||||||||||||||||||||||||| |||||||| |||||||||
Db    923 TTAGAAAGGGAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGGT  982

Qy   2334 AATATATTTTAGAAATATATATCAATAAATTCAATTTTAATCAATTAT 2393
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    983 AATATATTTTAGAAATATATATCAATAAATTCAATTTTAATCAATTAT 1042

Qy   2394 TAATTTTTTTTTTCTTTTTAAAGAGAAGCAGCTTTCTAGACTTTAAAGAATAAAGTCT 2453
          |||||||||||| ||||||||||||| :|| ||||| ||||||||||||||||||||
Db   1043 TAATTTTTTTTTCTTTTTAAAGAGRAAGCMGGCTTTTCTAGACTTTAAAGAATAAAGTCT 1102

Qy   2454 TTGGGAGGTCTCACGGTGTAGAGAGAGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGA 2513
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1103 TTGGGAGGTCTCACGGTGTAGAGAGAGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGA 1162

Qy   2514 GGGAAATCTCGTCGAAGGACACT 2537
          |||||||||||||||||||||||
Db   1163 GGGAAATCTCGTCGAAGGACACT 1186
```

```
RESULT 12
ABX73911
ID   ABX73911 standard; DNA; 1186 BP.
XX
AC   ABX73911;
XX
DT   18-MAR-2003   (first entry)
XX
DE   Human novel polynucleotide #739.
XX
KW   Human; gene; ds; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW   haemostatic; antiarteriosclerotic.
XX
OS   Homo sapiens.
XX
PN   US2002132753-A1.
XX
PD   19-SEP-2002.
XX
PF   17-JAN-2001; 2001US-00764864.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   28-JUN-2000; 2000US-0214886P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
```

```
Query Match          32.1%;  Score 1096.8;  DB 7;  Length 1186;
Best Local Similarity  98.4%;  Pred. No.1.3e-280;
Matches 1145;  Conservative  3;  Mismatches  12;  Indels  4;  Gaps  4;

Qy  1376  CCCAGGCTGTGGGCGCCATCATCGGGAAGAAGAGGGGGCACACATCAAACAGTGGCGAGAT  1435
                ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db    25  CCCACGGTCCGGCGCCATCATCGGGAAGAAGAGGGGGCACACATCAAACAGTGGCGAGAT    84

Qy  1436  TCGCCGGAGCCTCTATCAAGATTGCCCCTGCCGAAGGCCCAGCGTCAGCGAAAGGATGG  1495
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    85  TCGCGGGAGCCTCTATCAAGATTGCCCCTGCCGAAGGCCCAGCGTCAGCGAAAGGATGG   144

Qy  1496  TCATCATCACCGGCCACCGAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAAAC  1555
                ||||||||||||||  ||||||||| |||||||||||||||||||||||||||||||
Db   145  TCATCATCACCGGCCA-CGAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAAAC   203

Qy  1556  TGAAAGAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGAGTGGAAGCGCATATCAGAG  1615
                ||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||
Db   204  TGAAAGAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGGAACCCATTATCAGAG   263

Qy  1616  TGCCCTCCTTCCACAGCTGGCCGGGTGATTGGCAAGGTGCAAGACCGTGAACGAACTGC  1675
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   264  TGCCCTCCTTCCACAGCTGGCCGGGTGATTGGCAAGGTGCAAGACCGTGAACGAACTGC   323

Qy  1676  AGAACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGG  1735
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   324  AGAACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGG   383

Qy  1736  AAGTGATCGTCAGAATTATCGGGCACTTCTTGCTAGCCAGAGCAGAAATACCCTCAGGAGTGCCTCA  1795
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   384  AAGTGATCGTCAGAATTATCGGGCACTTCTTGCTAGCCAGAGCAGAAATACCCTCAGGAGTGCCTCA   443

Qy  1796  GGGAAATTG-TACAACAGGTGAAGCAGGAGCAGAGAGCAGAAATACCCTCAGGGAGTGCCTCA  1854
                |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   444  GGGAAATTGTTACAACAGGTGAAGCAGCAGAGAGCAGAAATACCCTCAGGGAGTCGCCTCA   503

Qy  1855  CAGCGCAGCAAGTGAGGCTCCCACAGGCACCACCAGGCACCAGGGAATGTAGCCCTTC  1914
                ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db   504  CAGCGCA-NAAGTGAGGCTCCCACAGGCACCACCAGGCACCAGGGAATGTAGCCCTTC   562

Qy  1915  CAACACCTGACAGAATGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAGACCA  1974
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   563  CAACACCTGACAGAATGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAGACCA   622

Qy  1975  TCTGAGGAATGAGAAGTCTGCGGAGGCGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCC  2034
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   623  TCTGAGGAATGAGAAGTCTGCGGAGGCGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCC   682

Qy  2035  AGGGGCGGAGGAGGGGCGGGGAAGGTCAGCGCAGGTTTGCCAGAACGCCACCGGCCCCGCCT  2094
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   683  AGGGGCGGAGGAGGGGCGGGGAAGGTCAGCGCAGGTTTGCCAGAACGCCACCGAGCCCGCCT   742

Qy  2095  CCCGCCCCCCAGGGGCTTCTGCAGCCAT-CCACTTCACCATCCACTTCACCATCCGATCTC  2153
                ||||||||||||||||||||||||||  ||||||||||||||  ||||||||||||||
Db   743  CCCGCCCCCCAGGGGCTTCTGCAGCCTTCAGCNATNCCACTTCACCATCCACTTCCGATCTC   802

Qy  2154  TCCTGAACTCCCAGCACGCTATCCTTTAGTGAACTAACATAGGTGAACGTGTTCAA  2213
                |||||||||||||||||  |||||||| |||||||||||||||||||||||||||
Db   803  TCCTGAACTCCCAGCACGCTATCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAA   862

Qy  2214  GCCAAGCAAAATGCACACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATA  2273
                ||||||||||||||||||||||  |||||||||||||||||||||||||||||||
Db   863  GCCAAGCAAAATGCACACCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATA   922

Qy  2274  TTAGAAGGGAAGATGTTAAGATAGTGCCTGTGGTTACACAGGGTGCCTGCAGCCGGT  2333
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   923  TTAGAAGGGAAGATGTTAAGATAGTGCCTGTGGGTTACACAGGGTGCCTGCAGCCGGT   982

Qy  2334  AATATATTTTAGAATAATATATCAAATAACTCAACTAACTCCAATTTTTAATCAATTAT  2393
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   983  AATATATTTTAGAATAATATATCAAATAACTCAACTAACTCCAATTTTTAATCAATTAT  1042
```

PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0226868P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0235327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.

XX
PA (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.

XX
PI Rosen CA, Ruben SM, Barash SC;

XX
DR WPI; 2003-147444/14.
DR P-PSDB; ABU55651.

XX
PT New polypeptides and nucleic acids, useful in gene therapy for treating,
PT inhibiting or preventing e.g. neural, immune system, muscular,
PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT renal disorders.

XX
PS Claim 1; SEQ ID NO 749; 402pp; English.

XX
CC The invention relates to human novel polypeptides and their associated
CC polynucleotides. The polypeptides and polynucleotides are useful in gene
CC therapy for treating, inhibiting or preventing neural disorders, immune
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent
CC human novel polynucleotides of the invention

XX
SQ Sequence 1186 BP; 338 A; 310 C; 302 G; 230 T; 0 U; 6 Other;

```
Qy     2394  TAATTTTTTTCTTTTTAAAGAGAAAGCAGGCTTTCTGAGACTTTAAAGAATAAAGTCT  2453
             |||||||||||||||||||||||||   ||||||   ||:||||:||||||||||||||
Db     1043  TAATTTTTTTCTTTTTAAAGAGRAAGCMGGCTTTCTGAGACTTTAAAGAATAAAGTCT  1102

Qy     2454  TTGGGAGGTCTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGA  2513
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1103  TTGGGAGGTCTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGA  1162

Qy     2514  GGGAAATCTCGTCGGAAGGACACT  2537
             ||||||||||||||||||||||||
Db     1163  GGGAAATCTCGTCGGAAGGACACT  1186
```

RESULT 13
AAS26151
ID   AAS26151 standard; cDNA; 822 BP.
XX
AC   AAS26151;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human cDNA encoding a novel secreted protein, Seq ID 330.
XX
KW   Human; immunosuppressive; antiarthritic; ss; antirheumatic; cytostatic;
KW   cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW   antibacterial; virucide; fungicide; opthalmalogical; vulnerary;
KW   secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; skin ageing; food additive;
KW   preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO20015322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001341.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216447P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   29-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   02-OCT-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.

```
PR    17-NOV-2000;  2000US-0242208P.
PR    17-NOV-2000;  2000US-0249209P.
PR    17-NOV-2000;  2000US-0249210P.
PR    17-NOV-2000;  2000US-0249211P.
PR    17-NOV-2000;  2000US-0249212P.
PR    17-NOV-2000;  2000US-0249213P.
PR    17-NOV-2000;  2000US-0249214P.
PR    17-NOV-2000;  2000US-0249215P.
PR    17-NOV-2000;  2000US-0249216P.
PR    17-NOV-2000;  2000US-0249217P.
PR    17-NOV-2000;  2000US-0249218P.
PR    17-NOV-2000;  2000US-0249244P.
PR    17-NOV-2000;  2000US-0249245P.
PR    17-NOV-2000;  2000US-0249264P.
PR    17-NOV-2000;  2000US-0249265P.
PR    17-NOV-2000;  2000US-0249297P.
PR    17-NOV-2000;  2000US-0249299P.
PR    17-NOV-2000;  2000US-0249300P.
PR    01-DEC-2000;  2000US-0250160P.
PR    01-DEC-2000;  2000US-0250391P.
PR    05-DEC-2000;  2000US-0251030P.
PR    05-DEC-2000;  2000US-0251988P.
PR    05-DEC-2000;  2000US-0256719P.
PR    06-DEC-2000;  2000US-0251479P.
PR    08-DEC-2000;  2000US-0251856P.
PR    08-DEC-2000;  2000US-0251868P.
PR    08-DEC-2000;  2000US-0251869P.
PR    08-DEC-2000;  2000US-0251989P.
PR    08-DEC-2000;  2000US-0251990P.
PR    11-DEC-2000;  2000US-0254097P.
PR    05-JAN-2001;  2001US-0256678P.
PA    (HUMA-) HUMAN GENOME SCI INC.
XX
PI    Rosen CA,  Barash SC,  Ruben SM;
XX
DR    WPI; 2001-488783/53.
DR    P-PSDB; AAU16164.
XX
PT    New nucleic acid molecules encoding 461 human secreted proteins for
PT    diagnosing, preventing, treating or ameliorating medical conditions and
PT    used as food additives or preservatives.
XX
PS    Claim 1; SEQ ID NO 330; 980pp; English.
XX
CC    The invention relates to isolated nucleic acid molecules and their
CC    encoded secreted proteins. The nucleic acids and proteins are used to
CC    prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC    rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC    in diagnosing a pathological condition or susceptibility to a
CC    pathological condition. Antibodies to the proteins can also be used in
CC    alleviating symptoms associated with the disorders and in diagnostic
CC    immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC    (ELISA). Disorders which are diagnosed or treated include autoimmune
CC    diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC    neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC    arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC    nervous system disorders e.g. Alzheimer's disease, infections caused by
CC    bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC    and many other disorders listed in the specification. The polypeptides
CC    can also be used to aid wound healing and epithelial cell proliferation,
CC    to prevent skin aging due to sunburn, to maintain organs before
CC    transplantation, for supporting cell culture of primary tissues, to
CC    regenerate tissues and in chemotaxis. The polypeptides can also be used
CC    as a food additive or preservative to increase or decrease storage
CC    capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC    minerals, cofactors and other nutritional components. The present
CC    sequence encodes a novel secreted protein of the invention. Note: The
CC    sequence data for this patent did not form part of the printed
```

```
Qy    779  CTGCAGAGAAGCCTGTCACCATCCATGCACCCCAGAGGGGGACTTCTGAAGCATGCCGCA    838
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  CTGCAGAGAAGCCTGTCACCATCCATGCACCCCAGAGGGGGACTTCTGAAGCATGCCGCA     60

Qy    839  TGATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTC    898
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  TGATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTC    120

Qy    899  TGAAAATCTTGGCACACAATGCTTGGTTGGAAGACTGATTGGAAAGAAGAAGGCAGAAATT    958
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  TGAAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAGAAGAAGGCAGAAATT   180

Qy    959  TGAAGAAAATTGAACATGAAACACGGGACCAAGATAACAATCCATCTTGCAGGATTTGA   1018
           ||||||||||||||:|||||:||||||||||||::||||||||||::||||:||||||||
Db    181  TGAAGAAAATTGAACATGAAACACAGGGACCAAGRTAACAATCCATCTTTGCAGGATTTGA   240

Qy   1019  GCATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTG   1078
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  GCATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTG   300

Qy   1079  CTGAGATAGAGATTATGAAGAAGCTCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTA   1138
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  CTGAGATAGAGATTATGAAGAAGCTCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTA   360

Qy   1139  ACCAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAG   1198
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  ACCAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAG   420

Qy   1199  GACTGTCCGTCTATCTCCACCAGCAGCAGGCGCCCGCCGCGGAGTCCCCCGCGTGCCCCCTACC   1258
           ||||||||||||||||||||||||||||||||||::||||||||||||:|:||||||||||
Db    421  GACTGTCCGTCTGTCTATCTCCANCAGCAGCAGGCGCCCGCNGGAG-TCCCCCGCTGCCCCCTACS   479

Qy   1259  ACCCCTTCACTACCTCCGGATACTTCTCCAGCTGTACCCCATCACCAGTTTGGCC   1318
           |:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    480  ASCCCTTCACTAMCCAMTCCGGATAWTTCTCCAGCTGTACCCCATCACCAGTTTGGCC   539

Qy   1319  CGTTCCCGCCATCATCACTCTTATCCAGACAGGAGATTGGAATCTTCATCCCAACCC   1378
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    540  CGTTCCCGCCATCATCACTCTTATCCAGACAGGAGATTGTGAATCTTCATCCCAACCC   599

Qy   1379  AGGCTGTGGGGCCATCATCGGGAAGAAGGGGCACACATCAAACAGCTGGCGAGATTCG   1438
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    600  AGGCTGTGGGGCCATCATCGGGAAGAAGGGGCACACATCAAACAGCTCGGCGAGATTCG   659

Qy   1439  CCGGAGCCTCTATCAAGATTGCCCCTGCGGAAGGGCCCAGGACGTCAGCGAAAGGAGGTCA   1498
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||::||||
Db    660  CCGGAGCCTCTATCAAGATTGCCCCTGCGGAAGGGCCCAGACGTCAGCGAAAGGATGGTCA   719

Qy   1499  TCATCACCGGGCCACCGGAA-GCCCAGTTCAAGGCCCAGGACGACGGATCTTTGGGAAACTCG   1557
           ||||||||||||||||||||:|:||||||||||||||||||:||:|||||||||::||||
Db    720  TCATCAMCGGGCCCAMCGGAATGCCCCAGTTCAAGGCCCAGGTTCAAGGCCCAGGACGGATYTTTGGGAAAYTG   779

Qy   1558  AAAG-AGGAAAACTTCTTTAACCCCAAAGAAGAAGAAGTGAAGCTG   1599
           |||| ||||||||||||||||||||||||||||||||||||||||||
Db    780  AAAGAAGGAAAATTTTTAACCCCAAAGAAGAAGAAGTGAAGCTG   822
```

```
RESULT 14
ABX73492
ID   ABX73492 standard; DNA; 822 BP.
XX
AC   ABX73492;
XX
DT   18-MAR-2003  (first entry)
XX
DE   Human novel polynucleotide #320.
XX
KW   Human; gene; ds; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
```

```
Query Match      22.7%;  Score 776.2;  DB 4;  Length 822;
Best Local Similarity  97.8%;  Pred. No. 1.9e-195;
Matches  805;  Conservative  10;  Mismatches  5;  Indels  3;  Gaps  3;
```

KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW   haemostatic; antiarteriosclerotic.
XX
OS   Homo sapiens.
XX
PN   US2002132753-A1.
XX
PD   19-SEP-2002.
XX
PF   17-JAN-2001; 2001US-00764864.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   28-JUN-2000; 2000US-0214886P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0254447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   22-AUG-2000; 2000US-0268868P.
PR   30-AUG-2000; 2000US-0229924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   08-SEP-2000; 2000US-0231413P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0244997P.
PR   27-SEP-2000; 2000US-0235834P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241809P.
PR   01-NOV-2000; 2000US-0244617P.
PR   17-NOV-2000; 2000US-0249299P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA, Ruben SM, Barash SC;
XX
DR   WPI; 2003-147444/14.
DR   P-PSDB; ABU55232.
XX
PT   New polypeptides and nucleic acids, useful in gene therapy for treating,
PT   inhibiting or preventing e.g. neural, immune system, muscular,

---

PT   respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT   renal disorders.
XX
PS   Claim 1; SEQ ID NO 330; English.
XX
CC   The invention relates to human novel polypeptides and their associated
CC   polynucleotides. The polypeptides and polynucleotides are useful in gene
CC   therapy for treating, inhibiting or preventing neural disorders, immune
CC   system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC   and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC   nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC   gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC   (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC   heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC   renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC   leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC   appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC   related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC   infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent
CC   human novel polynucleotides of the invention
XX
SQ   Sequence 822 BP; 234 A; 212 C; 201 G; 163 T; 0 U; 12 Other;

Query Match   22.7%;   Score 776.2; DB 7;   Length 822;
Best Local Similarity   97.8%;   Pred. No.1.9e-195;
Matches   805;   Conservative   10;   Mismatches   5;   Indels   3;   Gaps   3;

```
Qy    779 CTGCAGAGAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCA   838
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CTGCAGAGAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCA    60

Qy    839 TGATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAAGAGATTCTC   898
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAAGAGATTCTC   120

Qy    899 TGAAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAGGCAGAAATT   958
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGAAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAGGCAGAAATT   180

Qy    959 TGAAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCCATCTTTGCAGGATTTGA   1018
          ||||||||||||||||||||||||||||||||||:|::|||||||||||||||||||
Db    181 TGAAGAAAATTGAACATGAAACAGGGACCAAGRTAACAATCTCATCTTTGCAGGATTTGA   240

Qy   1019 GCATATACAACCCGGAAGAACCATCACTGTGAAGGGCACAGTTGAGCGCTGAGCGTGTG   1078
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCATATACAACCCGGAAGAACCATCACTGTGAAGGGCACAGTTGAGCGCTGAGCGTGTG   300

Qy   1079 CTGAGATAGAGATTATGAAGAAGTCGCGTGAGCTTTGAAAATGATATGCTGGCTGTTA   1138
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTGAGATAGAGATTATGAAGAAGTCGCGTGAGCTTTGAAAATGATATGCTGGCTGTTA   360

Qy   1139 ACCAACAAGCCAATCTGATCCCAGGGTTGAACCTTCAGCGCACTTGGCATCTTTTCAACAG   1198
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ACCAACAAGCCAATCTGATCCCAGGGTTGAACCTTCAGCGCACTTGGCATCTTTTCAACAG   420

Qy   1199 GACTGTCCGTGCTATCTCCACCAGCAGGGCCCCGGAGCTCCCCGCTGCCCCCTACC   1258
          ||||||||||||||||||||||||||||||||::|||||||||||||||||||||
Db    421 GACTGTCCGTGCTATCTCCANCAGCAGGGCCCCGNGGAG-TCCCCCGTGCCCCCTACS   479

Qy   1259 ACCCCTTCACTACCCACTCCGGATACTTCTCCAGCAGCAGAGATTGAATCTCTTCATCCAACCC   1318
          |::|||:||:||::|||||||::::||||||||||||||||||||||||||||||
Db    480 ASCCCTTCACTAMCCAMTCCGGATAMTTCTCCAGCAGCAGAGATTGTGAATCTCTTCATCCCAACCC   539

Qy   1319 CGTTCCCGCATCATCACTCTTATCCAGCAGCAGAGATTGTGAATCTTCTTCATCCCAACCC   1378
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    540 CGTTCCCGCATCATCACTCTTATCCAGCAGCAGAGATTGTGAATCTTCTTCATCCCAACCC   599

Qy   1379 AGGCTGTGGGCGCCATCATCGGGAAGAAGGGGCACACATCAAACAGTCTGGCAGATTCG   1438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    600 AGGCTGTGGGCGCCATCATCGGGAAGAAGGGGCACACATCAAACAGTCTGGCAGATTCG   659

Qy   1439 CCGGAGCCCTCTATCAAGATTGCCCCTGCGAAGGCGTCAGCGACGTCAGCGAAAGGATGGTCA   1498
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    660 CCGGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCAGACGTCAGCGAAAGATGGTCA 719

Qy   1499 TCATCACCGGCGCCACCGGAA-GCCCAGTTCAAGGCCCAGGACGGATCTTTGGGAAACTG 1557
          |||||:||||||||||| ||||  ::|||||||||||||||||||||:||||:|||||||
Db    720 TCATCAMCGGCCAMCGGATGCCCAGTTCAAGGCCCAGGACGGACGGATYTTTGGGAAAVTG 779

Qy   1558 AAAG-AGGAAAAACTTCTTTAACCCCAAAGAAGAGAAGTGAAGCTG 1599
          ||||  |||||||| ||||| |||||||||||||||||||||||||
Db    780 AAAGAAGGAAAAATTTTTTAAACCCCAAAGAAGAAGAAGTGAAGCTG 822
```

```
RESULT 15
AAC66035
ID   AAC66035    standard; cDNA; 1740 BP.
XX
AC   AAC66035;
XX
DT   21-FEB-2001  (first entry)
XX
DE   Human lung cancer-associated cDNA antigen L523S.
XX
KW   Lung cancer; therapy; human; tumor; immunogenic; cytostatic;
KW   vaccine; detection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200061612-A2.
XX
PD   19-OCT-2000.
XX
PF   03-APR-2000; 2000WO-US008896.
XX
PR   02-APR-1999;  99US-00285479.
PR   17-DEC-1999;  99US-00466396.
PR   30-DEC-1999;  99US-00476496.
PR   10-JAN-2000;  2000US-00480884.
PR   22-FEB-2000;  2000US-00510376.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T, Fan L;
XX
DR   WPI; 2000-628399/60.
DR   P-PSDB; AAB11365.
XX
PT   Isolated polypeptide comprising an immunogenic portion of a lung tumor
PT   protein is used for detecting and monitoring progression of lung cancer
PT   in a patient.
XX
PS   Claim 1a; Page 258-259; 261pp; English.
XX
CC   This invention describes a novel isolated polypeptide (I) which
CC   comprising an immunogenic portion of a lung tumor protein or variant (P2)
CC   which have cytostatic activity. The polypeptides and polynucleotides are
CC   used in compositions and vaccines to inhibit the development of cancer,
CC   especially lung cancer, in a patient. Methods described in the invention
CC   can be used to monitor the progression of a cancer by carrying out the
CC   detection at subsequent time points and comparing the results from the
CC   different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
CC   are treated with P2, polynucleotides encoding P2 or antigen presenting
CC   cells expressing P2 and then administered to the patient to inhibit
CC   development of cancer
XX
SQ   Sequence 1740 BP; 526 A; 406 C; 417 G; 391 T; 0 U; 0 Other;
```

```
Query Match          20.9%;   Score 712;   DB 3;   Length 1740;
Best Local Similarity  64.5%;   Pred. No. 3.7e-178;
Matches 1156;   Conservative     0;   Mismatches 555;   Indels 81;   Gaps 3;

Qy    73 ATGAACAAGCTTTACATCGGGACACCTTGAGCCCCGCCGTCACCGCCGACACCTCCGGCAG 132
         |||||||||||| ||||||||| ||||| |||  |||||  |||||  ||||||||||||
Db     1 ATGAACAAACTGTATATCGGAACCTCAGCCGAGAACGCCCGGAGAACGCGCCCCACCTCA 60

Db   1204 TCCGTGCTATCTCCACCAGCAGCAGGGCCCCGGGAGCTCCCCCGCTGCCCCCTACCACCCC 1263
```

Right column alignment sequences:

```
Db    133 CTCTTTGGGGACGAGGAAGCTGCCCCTGGCGGGACAGCAGAGTCCTGCTGAAGTCCGGCTACGCC 192

Qy     61 ATCTTCAAGGACGCCAAGATCTTCCTGTGTCGGGGACCCTTCCTGGTGAAGACTGCTCACGCG 120
          |||| ||||||||||||||||:||||||||||:||||||||||:|||||||||:||||||||
Db    193 TTCGTGGACTACCCCGACCAGAACTGGGCCCATCCGCGGCCATCGAGACCCCTTCGGGTAAA 252

Qy    121 TTCGTGGACTGCCCGACGAGAGCTGGGCCCTTCAAGGCCATCGAGGCGCTTTTCAGTGTAAA 180
          ||||||||||||||| ||||||||||||||||||||||||||||||:||||||:|||||||
Db    253 GTGGAATTGCATCGGAGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGTAAGGAGC 312

Qy    181 AATAGAACTGCACGGAACCATAGAAGTTGAGCACTCCGGTCCAAAAAAGGCAAAGGATT 240
          ||||||||||||||||||||| ||||||||||||||||:||||| ||||||||||||
Db    313 AGGAAAAATTCAGATTCGAAACATCCCTCCTACCTGCAGTGGGAGGTGTTGGATGGACTT 372

Qy    241 CGGAAAACTTCAGATACCGAAAATATCCCGCCTCATTTACAGTGGGAGGTGCTCGATAGTTTA 300
          ||||| ||||||||:|||||||||||||| |||:||||||||:||||||||||:||||||
Db    373 TTGGCTCAATATGGGACAGTGGAGAATGTGAACAAGTGAACAGTCAACACAGACAGAAACCGCC 432

Qy    301 CTAGTCCAGTATGGAGCTGTGAGAGCTGGTGAGCAAGTGAACAGTCAACACTGACTCGGAAACTGCA 360
          ||||||||||||| |||||||||||||||||:||||||||||||||||||:||||||||||
Db    433 GTTGTCAACGTCACATATGCAACAAGAGAAGAAGCAAAATAGCCATGGAGAAGCTAAGC 492

Qy    361 GTTGTAAATGTAACCTATTCCAGTAAGGACCAAGCTAGACACAGCACTAGACAAACTGAAT 420
          |||||| |||||||| ||||| |||:|||||||| |||:||||||||:||||| ||||||||
Db    493 GGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCCGATGAAGAGGTGAGC 552

Qy    421 GGATTTCAGTTAGAGAATTTCACCTTGAAAGTAGCCTATATCCCTGATGAAACGGCCGCC 480
          |||||||||||| |||| ||||||||||||||||||||||||||:||||||:||||| |||
Db    553 TCCCCTTCGCCCCCTCAGCGAGCCCCAGCGTCGGGGACCCACTCTTCCCGGGAGCAAGGC--- 609

Qy    481 CAGCAAAACCCCTTGCAGCAGCCCAGGTCGGGGCCCGGGAGGGGGCAGAGGGGGCTCCTCA 540
          ||||||||| |||| |||||||||||||| |||| |||||||||||| |||  :
Db    610 ------CACGGCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCGGTCGCGGATC 663

Qy    541 AGGCAGGGGTCTTCCAGATCCGTATCCAAGCAGAGAAACCATGTGATTTGCCTCGCGCCTG 600
          |||||||||||||:|||||||||||||||||||||||||||||||:|||||||||||||
Db    664 CTGGTCCCCACCAGTTGTTGGTCGCCATCATCGGAAAAGGAGGGCTTGACCATAAAAGAAC 723

Qy    601 CTGGTTCCCACCCAATTTGTTGGAGCCATCATCGGAAAAGAAGGGTGCCACCATTCGGAAC 660
          ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db    724 ATCACTAAGACAGACCCAGTCCAGGTAGATATCCAAGATATAAAATTCACAGAGAACTCTGGAGCTGCA 783

Qy    661 ATCACCAAACAGACCCAGTCCAGTCCAAAATCGATGTCCACCGTAAAGAAAAATGCGGGGCTGCT 720
          ||||| ||| ||||||||||||||| ||||||||||| ||||||||||| ||||| ||  ||||
Db    784 GAGAAGCCTGTCACCATCCATGCCACCCCAGAGAGGGACTTCTGAAGCATGCCGAAGATGATT 843

Qy    721 GAGAAGTCGATTACTATCCTCTACTCCTGAAGGCACCTCTGCGGCTTGTAAGTCTATT 780
          |||||| |||| |||:| ||:|:|| |||||| |:||||||||||| ||||  :||||
Db    844 CTTGAAATCATGCAGAAGAGGCAGATGAGCAAACTAGCCGAAGAGATTCCTCTGAAA 903

Qy    781 CTGGAGATTATGCATCATAAGGAGGAAGCTCAAGATATAAATTCACAGAGAGATCCCCTTGAAG 840
          |||||||| |||||| || |||||||||||:||||| || ||||||||||| ||||||||:
Db    904 ATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTGAAG 963

Qy    841 ATTTTAGCTCATAATAACTTGTTGGACGTCTTATTGGTAAAGAAGAGGAAGAAATCTTAA 900
          |||||||||||||| |||:| |||| || |||||||||||||||||||||||||||||||
Db    964 AAAATTGAACATGAAGACAGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATA 1023

Qy    901 AAAATTGAGACAGAGCAACTAAAATCACGATATCTCCATTGCAGGAATTGACGCTG 960
          ||||||||| ||||| ||||| ||||||||:|||  ||:||||||| ||||| |||
Db   1024 TACAACCCGGAAAGAACCACTATTACAGTAGTTAAAGGCACAGTTGAGCGGCGTGCTGAG 1083

Qy    961 TATAATCCAGAACCGCACTATTACAGTTAAAGGCAATGTTGAGACATGTGCCAAGCTGAG 1020
          ||| ||||||| ||||||||||||||| |||||||||||||||:||||||||||||||||
Db   1084 ATAGATTATGAGAAGAGTGCGTGAGGCCTTTGAAAATGATATGCTCGGCTGTTAACCAA 1143

Qy   1021 GAGGAGATCATGAAGAAAATCAGGAGTCTTATGGAAATGATATTGCTTCTATGAAATCTT 1080
          |||||||:||||||||||||:|||| ||:||:|||  |||||| |||:|| |||||:||
Db   1144 CAAGCAATCGATCCCAGGTTGAACCTCAGCGCAGCCTCACCGCACCGACTCCGGCAG 1203

Qy   1081 CAAGCACATTTAATTCCTGGATTAAATCTGAACCGTTGGGTCTGTTCCCACCCACTTCA 1140
          || ||||| ||  | |||||| || ||||||||||||||:|||||||||||||||:||
```

```
Db   1141  GGGATGCCACCTCCCACCTCAGGGGCCCCCTTCAGCCATGACTCCT------------ 1185
                  |||    |||||     |  |||||||||||    ||
Qy   1264  TTCACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTGGCCCGTTC 1323
                                              ||| | | |||||||  |
Db   1186  ------------------------------CCCTACCCGCAGTTTGAGC---------- 1204

Qy   1324  CCGCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTTCATCCCAACCCAGGCT 1383
                                             |     |  |||||| ||    |||
Db   1205  ----------AATCAGAAACGGAGACTGTTCATCTGTTTATCCCAGCTCTATCA 1248

Qy   1384  GTGGGCGCCATCATCGGGAAGAAGGGGCACACATCAAACAGCTGGCGAGATTCGCCGGA 1443
                  || | ||||||||  |||||||      |||||||||||||  ||||||||
Db   1249  GTCGGTGCCATCATCGGCAAGCAGGGCCAAGCACATCAAGCAGCTTTCTCGCTTTGCTGGA 1308

Qy   1444  GCCTCTATCAAGAATTGCCCTGCGGAAGGGCCCAGACGTCAGCGGAAAGGATGGTCATCATC 1503
                  |    | |||||||  |||||  |||    |||||||||||||||||||||
Db   1309  GCTTCAATTAAGAATTGCTCCAGCGGAAGCACCAGATGCTAAAGTGAGGATGGTGATTATC 1368

Qy   1504  ACCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTTGGGAAACTGAAAGAG 1563
                  |    |  |||||||||||||||  |||  ||||||  ||||| |||   |||||
Db   1369  ACTGGACCACCAGAGGCTCAGTTCAAGGCTCAGGGAAGAATTTATGGAAAAATTAAAGAA 1428

Qy   1564  GAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGGAAGCGCATATCAGAGTGCCCTCT 1623
                  |||||||  |  ||  |||||   || |||||||||  |||||  || ||||
Db   1429  GAAAACTTTGTTAGTCCTAAAGAAGAGGTGAAACTTGAAGCTCATATCAGAGTGCCCATCC 1488

Qy   1624  TCCACAGCTGCCGCGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTGCAGAACTTA 1683
                  |||    | ||  | | || | |||  | || |  | ||||| |||||||||
Db   1489  TTTGCTGCTGGCAGAGTTATTGGAAAAGAGGCAAAAACGGTGAATGAATCAGAATTTG 1548

Qy   1684  ACCAGTGCAGAAGTCATCGTGCCTGCTGACAAAGCCAGATGAAAATGAGGAAGTGATC 1743
                  ||||||| |||||| | || ||  ||||  ||||||   |||||||||| |||
Db   1549  TCAAGTGCAGAAGTTGTTGTCCCTCGTGACCCAGACACCTGATGAGAATGACCAAGTGGTT 1608

Qy   1744  GTCAGAATTATCGGGCACTTCTTTGCTCAGCCAGACTGCACAGCGCAAGATCAGGCAAATT 1803
                  |||||    |  |||  ||| || | |||| ||| | |   ||||||| ||||
Db   1609  GTCAAAATAACTGGTCACTTCTATGCTTGCCCAGTTGCCCAGAGAAAAATTCAGGAAATT 1668

Qy   1804  GTACACAGGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCAC 1855
                  |||||| |||| |  |   |    ||  | ||| ||| | ||  |
Db   1669  CTGACTCAGGTAAAGCACAGCACCAACAACAGAAGGCTCTGCAAAGTGGACCAC 1720
```

Search completed: July 21, 2004, 00:08:59
Job time : 1254 secs

```
                    Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        July 20, 2004, 23:48:10 ; Search time 226 Seconds
                                    (without alignments)
                                   8378.285 Million cell updates/sec

Title:         US-09-270-437D-6
Perfect score: 3412
Sequence:      1 ggcagcggaggaggcggaga........aacctgaaatgttattt 3412

Scoring table: IDENTITY NUC
               Gapop 10.0 , Gapext 1.0

Searched:      682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters:      1355418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     Issued_Patents_NA:*
               1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
               2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
               3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
               4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
               5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
               6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3411	100.0	3412	3	US-09-061-709-6	Sequence 6, Appli
2	3411	100.0	3412	4	US-09-899-651-6	Sequence 6, Appli
3	3143	92.1	3283	3	US-09-061-709-8	Sequence 8, Appli
4	3143	92.1	3283	4	US-09-899-651-8	Sequence 8, Appli
5	712	20.9	1740	4	US-09-643-597-347	Sequence 347, App
6	712	20.9	1740	4	US-09-542-615A-347	Sequence 347, App
7	712	20.9	1740	4	US-09-606-421B-347	Sequence 347, App
8	710.4	20.8	4159	3	US-09-061-709-4	Sequence 4, Appli
9	710.4	20.8	4159	4	US-09-899-651-4	Sequence 4, Appli
10	710.4	20.8	4181	4	US-09-643-597-175	Sequence 175, App
11	710.4	20.8	4181	4	US-09-480-884A-175	Sequence 175, App
12	710.4	20.8	4181	4	US-09-542-615A-175	Sequence 175, App
13	710.4	20.8	4181	4	US-09-606-421B-175	Sequence 175, App
14	691.8	20.3	2224	4	US-09-261-855-1	Sequence 1, Appli
15	570.6	16.7	1946	3	US-09-061-709-7	Sequence 7, Appli
16	570.6	16.7	1946	4	US-09-899-651-7	Sequence 7, Appli
17	489.2	14.3	1708	3	US-09-061-709-5	Sequence 5, Appli
18	489.2	14.3	1708	4	US-09-899-651-5	Sequence 5, Appli
c 19	223.8	6.6	317	4	US-09-621-976-13846	Sequence 13846, A
c 20	73.8	2.2	7218	1	US-08-232-463-14	Sequence 14, Appl
c 21	56.8	1.7	7218	4	US-08-232-463-14	Sequence 14, Appl
22	45	1.3	598	4	US-09-669-751-39	Sequence 39, Appl
23	44.6	1.3	4403765	3	US-09-103-840A-2	Sequence 2, Appli
24	44.6	1.3	4411529	3	US-09-103-840A-1	Sequence 1, Appli
25	44.4	1.3	5562	4	US-10-204-708-63	Sequence 63, Appli
c 26	44	1.3	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c 27	43.4	1.3	55298	4	US-09-491-356C-1	Sequence 1, Appli
c 28	42.8	1.3	289	3	US-09-007-005-17	Sequence 17, Appl
c 29	42.8	1.3	289	3	US-09-244-796-17	Sequence 17, Appl
c 30	42.8	1.3	17056	3	US-09-245-041-3	Sequence 3, Appli
c 31	42.2	1.2	80161	3	US-09-036-987A-1	Sequence 1, Appli
c 32	42.2	1.2	80161	3	US-09-370-700-1	Sequence 1, Appli
c 33	42.2	1.2	80161	4	US-09-603-207-1	Sequence 1, Appli
c 34	42	1.2	277	3	US-09-007-005-3	Sequence 3, Appli
c 35	42	1.2	832	4	US-09-244-796-3	Sequence 3, Appli
c 36	42	1.2	2813	4	US-09-621-976-2813	Sequence 2813, Ap
c 37	41.6	1.2	248	3	US-09-007-005-32	Sequence 32, Appl
c 38	41.6	1.2	248	3	US-09-244-796-32	Sequence 32, Appl
39	41.4	1.2	791	3	US-08-998-416-346	Sequence 346, App
40	41.4	1.2	7130	3	US-09-055-105-31	Sequence 31, Appl
41	41.2	1.2	362	3	US-09-018-584A-11	Sequence 11, Appl
42	41.2	1.2	1722	4	US-09-134-001C-1976	Sequence 1976, Ap
43	41.2	1.2	56520	4	US-09-338-907-179	Sequence 179, App
44	41.2	1.2	56520	4	US-09-218-207-179	Sequence 179, App
45	41	1.2	3138	1	US-07-867-106-4	Sequence 4, Appli

ALIGNMENTS

```
RESULT 1
US-09-061-709-6
; Sequence 6, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 6
; LENGTH: 3412
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-6

  Query Match             100.0%;  Score 3411;  DB 3;  Length 3412;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 3412;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy          1 GGCAGCGGAGGAGGCGGAGAGGCGCGGTACCGGCCGGGGAGCGCGGGGCTCGGGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGCAGCGGAGGAGGCGGAGAGGCGCGGTACCGGCCGGGGAGCGCGGGGCTCGGGG 60

Qy         61 AAGAGACGGATGATGAACAAGCTTTACATCGGGAGAACCTGAGCCCCGGTCACCGCCGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGAGACGGATGATGAACAAGCTTTACATCGGGAGAACCTGAGCCCCGGTCACCGCCGAC 120

Qy        121 GACCTTCCGGCAGCTCTTTGGGACAGGAGGAGCTGCCCCTGCGGGACAGGTCCTGCTGAAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GACCTTCCGGCAGCTCTTTGGGACAGGAGGAGCTGCCCCTGCGGGACAGGTCCTGCTGAAG 180

Qy        181 TCCGGCTACGGCCTTCGTGACTACCCGACCAGAACTGGGCCATCCGAGCACTGGAGACC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCGGCTACGGCCTTCGTGACTACCCGACCAGAACTGGGCCATCCGAGCACTGGAGACC 240

Qy        241 CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTACAGTTCTAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTACAGTTCTAA 300

Qy        301 AAGCTAAGGAGCAGGAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGGAGGTG 360
```

```
Db  1381 GCTGTGGGCGCCATCATCGGGAAGAAGAGGGGCACACATCAAACAGCTCGCCGAGATTCGCC 1440
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  1441 GGAGCCTCTATCAAGATGCCCTGCGGAAGGCCCAGCGTCAGCGGAAAGGATGGTCATC 1500
         ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db  1441 GGAGCCTCTATCAAGATGCCCTGCGGAAGGCCCAGCGTCAGCGGAAAGGATGGTCATC 1500

Qy  1501 ATCACCGGCCACCGGAAGCCCAGTTCAGGCCCGGACGGATCTTTGGGAAACTGAAA 1560
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 ATCACCGGGCCACCGGAAGCCCAGTTCAGGCCCGGACGGATCTTTGGGAAACTGAAA 1560

Qy  1561 GAGGAAACTTCTTAACCCAAAGAAGAAGTGAAGCGCATATCAGAGTGCCC 1620
         ||||||||||||||| |||||||||||||||||||||||||||||||||
Db  1561 GAGGAAACTTCTTAACCCAAAGAAGAAGTGAAGCTGGAAGCGCATATCAGAGTGCCC 1620

Qy  1621 TCTTCCACAGCTGCCGGGGTGATTGGCAAGGTGGCAAGACCGTGAACGAACTGCAGAAC 1680
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621 TCTTCCACAGCTGCCGGGGTGATTGGCAAGGTGGCAAGACCGTGAACGAACTGCAGAAAC 1680

Qy  1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCTGTGACCAAACGCCAGATGAAAATGAGGAAGTG 1740
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCTGTGACCAAACGCCAGATGAAAATGAGGAAGTG 1740

Qy  1741 ATCGTCAGAATTATCGGGCACTTCTTTGCTAGCAGCCAGATCGCACAGCGGAGTCGCCTCAGGGAA 1800
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741 ATCGTCAGAATTATCGGGCACTTCTTTGCTAGCAGCCAGATCGCACAGCGGAGTCGCCTCAGGGAA 1800

Qy  1801 ATTGTACAACAGGTGAAGCAGCAGCAGGCACCAGCAAAACAACGATGAATGTGAGCCTTCCAACAC 1860
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801 ATTGTACAACAGGTGAAGCAGCAGCAGGCACCAGCAAAACAACGATGAATGTGAGCCTTCCAACAC 1860

Qy  1861 AGCAAGTGAGGCTCCCACAGGACCAGCAAACAACGCAGCAAACCATCTGAG 1920
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861 AGCAAGTGAGGCTCCCACAGGACCAGCAAACAACGCAGCAAACCATCTGAG 1920

Qy  1921 CTGACAGAATGAGACCAAACGAGCCAGCCCTCGCCGAGGACTCTGCCGAGAACCCCAGGGGGC 1980
         ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921 CTGACAGAATGAGACCAAACGAGCCAGCCCTCGCCGAGGACTCTGCCGAGAACCCCAGGGGGC 1980

Qy  1981 GAATGAGAAGTCTGCGGAGGGCGGCGCCGGGGAGGTCAGCCAGGTTTGCCAGGGGC 2040
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981 GAATGAGAAGTCTGCGGAGGGCGGCGCCGGGGAGGTCAGCCAGGTTTGCCAGGGGC 2040

Qy  2041 CGAGGAGGGCGCGGGAGGTCAGCCAGTTGCCAGAACCACCAGCCCGCCTCCCGCC 2100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041 CGAGGAGGGCGCGGGAGGTCAGCCAGTTGCCAGAACCACCAGCCCGCCTCCCGCC 2100

Qy  2101 CCCCAGGGCTTCTGCAGCTTCAGCGCTTCAGTTGAGGCCTGTGCCAGTCTCCTGAA 2160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101 CCCCAGGGCTTCTGCAGCTTCAGCGCTTCAGTTGAGGCCTGTGCCAGTCTCCTGAA 2160

Qy  2161 CTCCCAGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161 CTCCCAGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2220

Qy  2221 AAAATGCACACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTTGTACATATTAGAAA 2280
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221 AAAATGCACACCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTTGTACATATTAGAAA 2280

Qy  2281 GGGAAGATGTTAAGATATGTGGCCTGTGGGGTTACACAGGGTGCCTGCAGCCGGTAATATAT 2340
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281 GGGAAGATGTTAAGATATGTGGCCTGTGGGGTTACACAGGGTGCCTGCAGCCGGTAATATAT 2340

Qy  2341 TTTAGAAATAATATCAAAATAACTAACTCCAATTTTAATCAATTATTAATTTT 2400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341 TTTAGAAATAATATATCAAAATAACTAACTCCAATTTTAATCAATTATTAATTTT 2400

Qy  2401 TTTTTCTTTTTAAAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGAG 2460
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401 TTTTTTCTTTTTAAAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGAG 2460

Qy  2461 GTCTCACGGTGTGAGAGGAGGAGCTTTGAGCGACCCACCCACCCGCACAAAATTCACCCAGAGGGAAAT 2520
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461 GTCTCACGGTGTGAGAGGAGGAGCTTTGAGCGACCCACCCACCCGCACAAAATTCACCCAGAGGGAAAT 2520
```

```
Db   301 AAGCTAAGGAGGCAGGAAAATTCAGATTCGAAACATCCCTCCTCCACCTGCAGTGGGAGGTG 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   361 TTGGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGAC 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TTGGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGAC 420

Qy   421 ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAAGGAAGAAGCAAAAATAGCCATG 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ACAGAAACCGCCGTGTCAACGTCACATATGCAACAAGAAGAAGCAAAAATAGCCATG 480

Qy   481 GAGAAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGAT 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GAGAAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGAT 540

Qy   541 GAAGAGGTGAGCTCCCCTTGCGCCCCTCAGCGAGCCCAGCCTGGGGACCACTCTTCCCGG 600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GAAGAGGTGAGCTCCCCTTGCGCCCCTCAGCGAGCCCAGCCTGGGGACCACTCTTCCCGG 600

Qy   601 GAGCCAAGGCCACCGCCCCTCTTCAGGCCGCAGACAGATTGATTCCCGTGCGG 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GAGCCAAGGCCACCGCCCCTCTTCAGGCCGCAGACAGATTGATTCCCGTGCGG 660

Qy   661 ATCCTGTCCCCACCAGTTTGTTGGTGCCATCATCGGGAAAGGAGGGCTTGACCATAAAG 720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 ATCCTGTCCCCACCAGTTTGTTGGTGCCATCATCGGGAAAGGAGGGCTTGACCATAAAG 720

Qy   721 AACATCACTAAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAAGAACTCTGGAGCT 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 AACATCACTAAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAAGAACTCTGGAGCT 780

Qy   781 GCAGAGAAGCCTGTCACCATCCATCCATGCCACCCCAGAGGGACTTCTGAAGCATGCCGCATG 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GCAGAGAAGCCTGTCACCATCCATCCATGCCACCCCAGAGGGACTTCTGAAGCATGCCGCATG 840

Qy   841 ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACTAGCCGAAGATTCCTCTG 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACTAGCCGAAGAGATTCCTCTG 900

Qy   901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAGAAGAAGGCAGAAATTTG 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAGAAGAAGGCAGAAATTTG 960

Qy   961 AAGAAAATTGAACATGAAACAGGGACCAAGATACAAATCTCATCTTTGCGAGGATTTGAGC 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCGAGGATTTGAGC 1020

Qy  1021 ATATACAACCCGGAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCT 1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 ATATACAACCCGGAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCT 1080

Qy  1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATAATGTGATGTGTTAAC 1140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATAATGTGATGTGTTAAC 1140

Qy  1141 CAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGA 1200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 CAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGA 1200

Qy  1201 CTGTCCGTGCTATCTCCACCACTCCGATACTCTTCCACAGGCCCCCGCGAGCTCCCCTACCAC 1260
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 CTGTCCGTGCTATCTCCACCACTCCGATACTCTTCCACAGGCCCCCGCGAGCTCCCCTACCAC 1260

Qy  1261 CCCTTCACTCACCACTCCGATATCTTATCCGACAGGAGGATTGTGAATCTTCTTCATCCCCAACCCAG 1320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 CCCTTCACTCACCACTCCGATATCTTATCCGACAGGAGGATTGTGAATCTTCTTCATCCCCAACCCAG 1320

Qy  1321 TTCCCGCATCATCATCATTATCCGACAGGAGGATTGTGAATCTTCATCCCCAACCCAG 1380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 TTCCCGCATCATCATCATTATCCGACAGGAGGATTGTGAATCTTCATCCCCAACCCAG 1380

Qy  1381 GCTGTGGGCGCCATCATCGGGAAGAAGAGGGGCCACACATCAAACAGCTCGCCGAGATTCGCC
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
;   APPLICANT: Old, Lloyd J.
;   TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
;   TITLE OF INVENTION: Associated
;   TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
;   FILE REFERENCE: LUD 5538
;   CURRENT APPLICATION NUMBER: US/09/899,651
;   CURRENT FILING DATE: 2001-07-06
;   PRIOR APPLICATION NUMBER: US/09/061,709
;   PRIOR FILING DATE: 1998-04-17
;   NUMBER OF SEQ ID NOS: 8
;   SEQ ID NO 6
;     LENGTH: 3412
;     TYPE: DNA
;     ORGANISM: Homo sapiens
;     FEATURE:
US-09-899-651-6

Query Match             100.0%;  Score 3411;  DB 4;  Length 3412;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 3412;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy       1 GGCAGCGGAGGAGGCGAGGAGCCCGGGGTACCGGCCGCGGGGAGCCGGCCGCGGGGGTTCGCGGGG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GGCAGCGGAGGAGGCGAGGAGCCCGGGGTACCGGCCGCGGGGAGCCGGCCGCGGGGGTTCGCGGGG 60

Qy      61 AAGAGACCGATGATGAACAAGCTTTACATCGGGAGAGTCGCCCCGCCGTCCACCGCGCCGAC 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AAGAGACCGATGATGAACAAGCTTTACATCGGGAGAGTCGCCCCGCCGTCCACCGCGCCGAC 120

Qy     121 GACCTCCGGCAGCTCTTTGGGGACAGAGAAGCTGCCCCGTCGGCGGGACAGTCCTGCTGAAG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GACCTCCGGCAGCTCTTTGGGGACAGAGAAGCTGCCCCGTCGGCGGGACAGTCCTGCTGAAG 180

Qy     181 TCCGGCTACGCCTTCGTCGTGACTACCCCGACTACCGGCCATCCGCGGCCATCGAGACC 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TCCGGCTACGCCTTCGTCGTGACTACCCCGACTACCGGCCATCCGCGGCCATCGAGACC 240

Qy     241 CTCTCGGTAAAGTGGAATTGCATGGGAAAATCATGGAAATCAGTTGATTACTCAGTCTCTAAA 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CTCTCGGTAAAGTGGAATTGCATGGGAAAATCATGGAAATCAGTTGATTACTCAGTCTCTAAA 300

Qy     301 AAGCTAAGAGCAGGAGAAATTCAGATTCGAAAACATCCCCTCCTCAGTGGACAGTCAACACAGAC 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AAGCTAAGAGCAGGAGAAATTCAGATTCGAAAACATCCCCTCCTCAGTGGACAGTCAACACAGAC 360

Qy     361 TTGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGGAACAGTCAACACAGAC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TTGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGGAACAGTCAACACAGAC 420

Qy     421 ACAGAAACCGCCGGTGCAAGCGTCACATATGCAACAAGAGAAGAAGCAAAAATAGCCATG 480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ACAGAAACCGCCGGTGCAAGCGTCACATATGCAACAAGAGAAGAAGCAAAAATAGCCATG 480

Qy     481 GAGAAGCTAAGCGGGCATCAGTTTGAGACTACTCCTTCAAGATTTCTACATCCCGGAT 540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GAGAAGCTAAGCGGGCATCAGTTTGAGACTACTCCTTCAAGATTTCTACATCCCGGAT 540

Qy     541 GAAGAGGTGAGCTCCCCTTCGCCCCTGGGGCACTTCTCAGGCGCCCAGCGTGGGGACCACTCTTCCCGG 600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GAAGAGGTGAGCTCCCCTTCGCCCCTGGGGCACTTCTCAGGCGCCCAGCGTGGGGACCACTCTTCCCGG 600

Qy     601 GAGCAAGGGCCACGCCACGCCCCTGGGGGCACTTCTCAGGCCAGCAGATTGATTTCCCGCTGCGG 660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GAGCAAGGGCCACGCCACGCCCCTGGGGGCACTTCTCAGGCCAGCAGATTGATTTCCCGCTGCGG 660

Qy     661 ATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGGAAAGGGCTTGACCATAAAG 720
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 ATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGGAAAGGGCTTGACCATAAAG 720

Qy     721 AACATCACTAAGCAGACCCAGTCCCGGTAGATATCCATAGAAAGAGAAACTCTGGAGCT 780
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 AACATCACTAAGCAGACCCAGTCCCGGTAGATATCCATAGAAAGAGAAACTCTGGAGCT 780
```

```
Qy    2521 CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGATGTCAACAGAGAAGGGATA 2580
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2521 CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGATGTCAACAGAGAAGGGATA 2580

Qy    2581 CCGTCTCCTGAAGAGGAAACTTCTGTCACTCCTCATGCCTCGTCTTAGCTCATACACCCATT 2640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2581 CCGTCTCCTGAAGAGGAAACTTCTGTCACTCCTCATGCCTCGTCTTAGCTCATACACCCATT 2640

Qy    2641 TCTCTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCCTCTGTCT 2700
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2641 TCTCTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCCTCTGTCT 2700

Qy    2701 CTTCTGTTTATCTCTCCCCTCCCCCCCTTCTTCTCCATCTCCCATTCTTTGAA 2760
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2701 CTTCTGTTTATCTCTCCCCTCCCCCCCTTCTTCTCCATCTCCCATTCTTTGAA 2760

Qy    2761 TTTCCTCATCCTCCATCTCAATCCGTATCTACGGCACCCCCCCCCCCCAGGCAAAGCA 2820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2761 TTTCCTCATCCTCCATCTCAATCCGTATCTACGGCACCCCCCCCCCCCAGGCAAAGCA 2820

Qy    2821 GTGCTCGAGTATCATCACACAAAAGGCAAACACACAAACCAGCCTCAA 2880
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db    2821 GTGCTCGAGTATCATCACACAAAAGGCAAACACACAAACCAGCCTCAA 2880

Qy    2881 CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGAAG 2940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2881 CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAG 2940

Qy    2941 AGGAAAAACAGGAACCCACCAAACCACCAATCACCAAAGAAAAAATTCCACAATG 3000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2941 AGGAAAAACAGGAACCCACCAAACCACCAATCACCAAAGAAAAAATTCCACAATG 3000

Qy    3001 AAAGAAATGTATTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATG 3060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3001 AAAGAAATGTATTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATG 3060

Qy    3061 ATTCCTTTCTTTAAAAAAAAAATGTGGAGGAAGTAGAAATTTACCAGGTTGTTGGCC 3120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3061 ATTCCTTTCTTTAAAAAAAAAATGTGGAGGAAGTAGAAATTTACCAGGTTGTTGGCC 3120

Qy    3121 CAGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAAACACAGAAGAAGCTACCTCAG 3180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3121 CAGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAAACACAGAAGAAGCTACCTCAG 3180

Qy    3181 GTGTTTTTACCTCAGCACCTGCCTTGGTTGTGTGTTCCCTTAGAGATTTTGTAAAGCTGATAG 3240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3181 GTGTTTTTACCTCAGCACCTGCCTTGGTTGTGTGTTCCCTTAGAGATTTTGTAAAGCTGATAG 3240

Qy    3241 TTGGAGCATTTTTTTATTTTTAATAAAAATGAGTTGGAAAAAATAAGAATATCAACT 3300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3241 TTGGAGCATTTTTTTATTTTTAATAAAAATGAGTTGGAAAAAATAAGAATATCAACT 3300

Qy    3301 GCCAGCCTGGAGAAGTGACAGTCCAAGTGTGCAACAGCTGTTCTTCCGCT 3360
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3301 GCCAGCCTGGAGAAGTGACAGTCCAAGTGTGCAACAGCTGTTCTTCCCGCT 3360

Qy    3361 AGCCAAGAACNATATGGCTTCTTTTGGACAAAACTTGAAAATGTTTATTT 3412
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3361 AGCCAAGAACNATATGGCTTCTTTTGGACAAAACTTGAAAATGTTTATTT 3412
```

```
RESULT 2
US-09-899-651-6
; Sequence 6, Application US/09899651
; Patent No. 6576756
; GENERAL INFORMATION:
;   APPLICANT: Chen, Yao-Tseng
;   APPLICANT: Gure, Ali
;   APPLICANT: Tsang, Solam
;   APPLICANT: Stockert, Elisabeth
;   APPLICANT: Jager, Elke
;   APPLICANT: Knuth, Alexander
```

```
Qy   781  GCAGGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATG  840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781  GCAGGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATG  840

Qy   841  ATTCTTGAAATCATGCGAGAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG  900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841  ATTCTTGAAATCATGCGAGAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG  900

Qy   901  AAAATCTTGGCACACAATGGCTTGGTTGGAAGAAAGAAGGCAGAAATTTG  960
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901  AAAATCTTGGCACACAATGGCTTGGTTGGAAGAAAGAAGGCAGAAATTTG  960

Qy   961  AAGAAAATTGAACATGAAACAACCGGAAAGAACCAAGATAACAATCTCATCTTTGCCAGGATTTGAGC  1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961  AAGAAAATTGAACATGAAACAACCGGAAAGAACCAAGATAACAATCTCATCTTTGCCAGGATTTGAGC  1020

Qy  1021  ATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCT  1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021  ATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCT  1080

Qy  1081  GAGATAGAGATTATGAAGAAGCTGCGTAGGCCTTTGAAAATGATGTCTGGCTGTGTTAAC  1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081  GAGATAGAGATTATGAAGAAGCTGCGTAGGCCTTTGAAAATGATGTCTGGCTGTGTTAAC  1140

Qy  1141  CAACAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGA  1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141  CAACAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGA  1200

Qy  1201  CTGTCCGTGTATCTCCACCACTGCTCCACCAGAGGGCCCCGGCGAGCTCCCCGCTGCCCCCTACCAC  1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201  CTGTCCGTGTATCTCCACCACTGCTCCACCAGAGGGCCCCGGCGAGCTCCCCGCTGCCCCCTACCAC  1260

Qy  1261  CCCTTCACTACACCCGGATACTTCTCCAGCTGTACCCCGTGTACCCCATCACCAGTTGGGCCCG  1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261  CCCTTCACTACACCCGGATACTTCTCCAGCTGTACCCCGTGTACCCCATCACCAGTTGGGCCCG  1320

Qy  1321  TTCCCGCATCATCACTCTTATCCGGAGAGAAGGGGGCACACATCAAACAGCTCGGCGAGATTGCC  1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321  TTCCCGCATCATCACTCTTATCCGGAGAGAAGGGGGCACACATCAAACAGCTCGGCGAGATTGCC  1380

Qy  1381  GCTGTGTGGGGGCCATCATCGGGAAGGGGCACACATCAAACAGCTCGGCGAAGGATTGGTCATC  1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381  GCTGTGTGGGGCGCCATCATCGGGAAGGGGCACACATCAAACAGCTCGGCGAAGGATTGGTCATC  1440

Qy  1441  GGAGCCTCATCAAGATTGCCCCGGAAGCCCAGTTCAAGGCCCAGGAGATCTTTGGGAAACTGAAA  1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441  GGAGCCTCATCAAGATTGCCCGGGAAGCCCAGTTCAAGGCCCAGGAGATCTTTGGGAAACTGAAA  1500

Qy  1501  ATCACCGGGCCACCCGAAGCCCAGTTCAAGGCCCAGGAGACGGATCTTTGGGAAACTGAAG  1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501  ATCACCGGGCCACCCGAAGCCCAGTTCAAGGCCCAGGAGACGGATCTTTGGGAAACTGAAG  1560

Qy  1561  GAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGCAAGATGAAGCTGGAAGCGCATATCAGAGTGCCC  1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561  GAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGCAAGATGAAGCTGGAAGCGCATATCAGAGTGCCC  1620

Qy  1621  TCTTCCACAGCTGCCGGGGTGATTGGCCAAGAGGTTGCAAGACCGTGCAAGACTGCAGAAC  1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621  TCTTCCACAGCTGCCGGGGTGATTGGCCAAGAGGTTGCAAGACCGTGCAAGACTGCAGAAC  1680

Qy  1681  TTAACCAGTTGCAACAGTTGCCTCGTGCCTCGTGCCCTGACCAAACGCCAGATGAAAATGAGGAAGTG  1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681  TTAACCAGTTGCAACAGTTGCCTCGTGCCTCGTGCCCTGACCAAACGCCAGATGAAAATGAGGAAGTG  1740

Qy  1741  ATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCGAGACTCCAGACTGCACAGCTGCACAGGGAA  1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741  ATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCGAGACTCCAGACTGCACAGCTGCACAGGGAA  1800

Qy  1801  ATTGTACAACAGGTTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGC  1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801  ATTGTACAACAGGTTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGC  1860

Qy  1861  AGCAAGTGAGGCTCCCAGGCACCAGCAACAACGGATGAATGTAGCCCTTCCAACAATG  1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861  AGCAAGTGAGGCTCCCAGGCACCAGCAACAACGGATGAATGTAGCCCTTCCACAC  1920
```

```
Qy  1921  CTGACAGAATGAGACCAAACGCCAGATGAGCCAGCCAGCCCAGATCGGGAGGCCTGAGCGCCCAGGGGC  1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921  CTGACAGAATGAGACCAAACGCCAGATGAGCCAGCCAGCCCAGATCGGGAGGCAAACCATCTGAG  1980

Qy  1981  GAATGAGAAGTCTGCGGAGGCGGCCCTGCGCCGAGGCTCGAGGCCTTGCCAGAACCACCGAGCCCCGCCTCCCGCC  2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981  GAATGAGAAGTCTGCGGAGGCGGCCCTGCGCCGAGGCTCGAGGCCTTGCCAGAACCACCGAGCCCCGCCTCCCGCC  2040

Qy  2041  CGGAGGAGGGCGGGGGAGGTCAGCCAGGTTTGCCAGAACCACCGAGCCCGCCTCCCGCC  2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041  CGGAGGAGGGCGGGGGAGGTCAGCCAGGTTTGCCAGAACCACCGAGCCCGCCTCCCGCC  2100

Qy  2101  CCCCAGGCCTTCAGCATCCATCCAGCTTCACTTCACCACTCCGATCTCTCTGAA  2160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101  CCCCAGGCCTTCGCAGGCTTCAGCATCCATCCAGCTTCACCATCCGGATCTCCTCTGAA  2160

Qy  2161  CTCCCAGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC  2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161  CTCCCAGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC  2220

Qy  2221  AAAATGACACACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAA  2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221  AAAATGACACACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAA  2280

Qy  2281  GGGAAGAGATGTTAAGATATGTGCCTGTGGGTTACACAGGGTGCCTGCAGCGTAATATAT  2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281  GGGAAGAGTGTTAAGATATGTGCCTGTGGGTTACACAGGGTGCCTGCAGCGTAATATAT  2340

Qy  2341  TTTAGAAATAATAATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTT  2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341  TTTAGAAATAATAATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTT  2400

Qy  2401  TTTTCTTTTTAAAGAGAAAGCAGCTTTCTAGACTTAAAGAATAAAGTCTTTGGGAG  2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401  TTTTCTTTTTAAAGAGAAAGCAGCTTTCTAGACTTAAAGAATAAAGTCTTTGGGAG  2460

Qy  2461  GTCTCACGGTGTGAGAGGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGAGGAGAAAT  2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461  GTCTCACGGTGTGAGAGGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGAGGAGAAAT  2520

Qy  2521  CCGTCTCCGAAGGACACTCACGGCAGTTCTGATCACCTGTGTATGTCAACGAGAGGGATA  2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521  CCGTCTCCGAAGGACACTCACGGCAGTTCTGATCACCTGTGTATGTCAACGAGAGGGATA  2580

Qy  2581  CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCACTCCTGTCGCTGTCTAGCTCTCATCACCATT  2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581  CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCACTCCTGTCGCTGTCTAGCTCTCATCACCATT  2640

Qy  2641  TCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCT  2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641  TCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCT  2700

Qy  2701  CTCTCGTGTTTATCTCCTCCAATCCCATCTCTTCTTCCCCCCCCCCATTCTCCATTCTTTTGAA  2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701  CTCTCGTGTTTATCTCCTCCAATCTCAATCTCTTCTTCCCCCCCCCCATTCTCCATTCTTTTGAA  2760

Qy  2761  TTTCCTCCATCCTCAATCCCGTATCTACGGCACCCCCCCCCCCCCCCCAGGCCAAAGCA  2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761  TTTCCTCCATCCTCCATCTCAATCCCGTATCTACGGCACCCCCCCCCCCCCCCCAGGCAAAGCA  2820

Qy  2821  GTGCTCTGAGTATCACATCACACAAAAGGACAAAAGAGAGTCAATGGTACTTGTCCTAGGGTTTTGGAAG  2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821  GTGCTCTGAGTATCACATCACACAAAAGGACAAAAGAGAGTCAATGGTACTTGTCCTAGGGTTTTGGAAG  2880

Qy  2881  CTTACACTTGGTTACTCAAAAGAATCAATGGTACTTGTCCTAGGCGTTTTGGAAG  2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881  CTTACACTTGGTTACTCAAAAGAATCAATGGTACTTGTCCTAGGCGTTTTGGAAG  2940

Qy  2941  AGGAAAACAGGAACCCACCAGCACCAGCAACAACCAATCAACCAAACAAAGAAAAATTCCACAATG  3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  2941 AGGAAAAACAGGAGACCCACCAGCAAACCAAATCAACCAAACAAGAGAAAAAATTCCACAATG 3000
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
         TCCGGCTACGCCTTCGTGGACTACCCGACCAGCAGTGGGCCATCCGCCGCGACTGAGACC 240
Qy   181

Qy  3001 AAAGAATGTATTTTGTCTTTTTGCATTTTGGTGTGTATAAGCCATCAATATTCAGCAAAATG 3060
         |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db  3001 AAAGAATGTATTTTGTCTTTTTGCATTTTGGTGTGTATAAGCCATCAATATTCAGCAAAATG 3060
         CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA 300
Db   241

Db  3061 ATTCCTTTCTTTAAAAAAAAAATGTGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCC 3120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3061 ATTCCTTTCTTTAAAAAAAAAATGTGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCC 3120
         AAGCTAAGGAGCAGGAGAAAATCAGATTGAAACATCCTCCTCCACCTGCAGTGGGACGTG 360
Qy   301

Qy  3121 CAGGGCGGTTAAATTTCACAGATTTTTTAACGAGAAAAACACACAGAGAAGAAGCTACCTCAG 3180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3121 CAGGGCGGTTAAATTTCACAGATTTTTTAACGAGAAAAACACACAGAGAAGAAGCTACCTCAG 3180
         TTGGATGGACTTTGGCTCAATATGGACAGTGGAGAATGTGAACAAGTCAACACAGAC 420
Db   361

Db  3181 GTGTTTTTACCTGACCACCTTGCTCTTGTGTTCTTGTGTTTTTCCCTTAGAGATTTTGTAAAGCTGATAG 3240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3181 GTGTTTTTACCTGACCACCTTGCTCTTGTGTTCTTGTGTTTTTCCCTTAGAGATTTTGTAAAGCTGATAG 3240
         ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAAGAAGAAGCAAAAATAGCCATG 480
Qy   421

Qy  3241 TTGGAGCATTTTTATTTTTAATAAAAATGAGTTGGAAAAAAATAAGATAATCAACT 3300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3241 TTGGAGCATTTTTATTTTTAATAAAAATGAGTTGGAAAAAAATAAGATAATCAACT 3300
         GAGAAGGTAAGCGGCCATCAGTTGAGAACTACTCCTCAAGATTTTCCTACATCCCGGAT 540
Db   481

Db  3301 GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCT 3360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3301 GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCT 3360
         GAAGAGGTGAGCTCCCCTTCGCCCCTCCAGCGTGGGACCACTCTTCCCGG 600
Qy   541

Qy  3361 AGCCAAGAACCNATATGGCCTTCTTTTGACAAACCTTGAAAATGTTTATTT 3412
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3361 AGCCAAGAACCNATATGGCCTTCTTTTGACAAACCTTGAAAATGTTTATTT 3412
         GAGCAAGGCCACGCCCGTGGGGCACTTCTCAGCCAGCAGACAGATTGATTCCCGCTGCGCG 660
Db   601

Qy       ATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCCGGAAAAGAGAGGGCTTGACCATAAAG 720
Db       ATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCCGGAAAAGAGAGGGCTTGACCATAAAG 720

Qy       AACATCACTAAGCACCAGTCCCGGTAGTATCCGGTAGATATCCATAGAAAGAACCTGGAGCT 780
Db       AACATCACTAAGCACCAGTCCCGGTAGTATCCGGTAGATATCCATAGAAAGAACCTGGAGCT 780

Qy       GCAGAAGAAGCCTGTCACCATCCATGCCACCCCAGAGAGGGACTTCTGAGCACGGCATGCCGCATG 840
Db       GCAGAAGAAGCCTGTCACCATCCATGCCACCCCAGAGAGGGACTTCTGAGCACGGCATGCCGCATG 840

Qy       ATTCTTGAAATCATGCAGAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG 900
Db       ATTCTTGAAATCATGCAGAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG 900

Qy       AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAAATTTG 960
Db       AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAAATTTG 960

Qy       AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020
Db       AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020

Qy  1021 ATATACAACCCGGAAAGAAGAACCATCACTGTGAAGGGCACAGTTGAAGGGCACAGTTGCCCAGTGCT 1080
Db  1021 ATATACAACCCGGAAAGAAGAACCATCACTGTGAAGGGCACAGTTGAAGGGCACAGTTGCCCAGTGCT 1080

Db  1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGAAGCCTTTGAAAATGATATGCTCGGCTGTTAAC 1140
Db  1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGAAGCCTTTGAAAATGATATGCTCGGCTGTTAAC 1140

Qy  1141 CAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCGACTCTTTCAACAGGA 1200

Db  1141 ------------------------------------------------------- 1141

Qy  1201 CTGTCCGTGCTATCTCCACCAGCAGAGGCCCGGCTCCCCGCTGCCCCCTACCAC 1260
Db  1141 ------------------------------------------------------- 1141

Qy  1261 CCCTTCACTACCCACCTCCGATACTTCTCCAGCCTGTACCCCAGTTGGCCCG 1320
```

RESULT 3
US-09-061-709-8
; Sequence 8, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 8
; LENGTH: 3283
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-8

Query Match   92.1%; Score 3143; DB 3; Length 3283;
Best Local Similarity 96.2%; Pred. No. 0;
Matches 3283; Conservative 0; Mismatches 0; Indels 129; Gaps 1;

```
Qy     1 GGCAGCGCGAGGAGGCGAGGAGCGGCCTACGGGCGCCCGGGTACGGGGGAGCCCGGCGTCCTCGGGGG 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GGCAGCGCGAGGAGGCGAGGAGCGGCCTACGGGCGCCCGGGTACGGGGGAGCCCGGCGTCCTCGGGGG 60

Qy    61 AAGAGACGGATGATGAACAAGCTTTACATCGGGAACAAGTTTGCCCGGCCGCGGAACCTGCCCCGGACC 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 AAGAGACGGATGATGAACAAGCTTTACATCGGGAACAAGTTTGCCCGGCCGCGGAACCTGCCCCGGACC 120

Qy   121 GACCTCCGGCCAGCTCTTTGGGGACAGGAAGCTGCCCTGGGACAGGTTCCTGTCTGAAG 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GACCTCCGGCCAGCTCTTTGGGGACAGGAAGCTGCCCTGGGACAGGTTCCTGTCTGAAG 180

Qy   181 TCCGGCTACGCCTTCGTGGACTACCCGACCAGCAGTGGGCCATCCGCCGCGACTGAGACC 240
```

```
Db   1141 ---------ACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGCCCG 1191

Qy   1321 TTCCCGCCATCATCACTCTTATCCGAGACGAGGAGATTGTGAATCTCTTCATCCCAACCCAG 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1192 TTCCCGCCATCATCACTCTTATCCGAGACGAGGAGATTGTGAATCTCTTCATCCCAACCCAG 1251

Qy   1381 GCTGTGTGGCGCCATCATCGGGAAGAAGAGGGGCCACACATCAAACAGCTGGCGAGATTCGCC 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1252 GCTGTGTGGCGCCATCATCGGGAAGAAGAGGGGCCACACATCAAACAGCTGGCGAGATTCGCC 1311

Qy   1441 GGAGCCTCTATCAAGAATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAGGATGTGTCATC 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1312 GGAGCCTCTATCAAGAATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAGGATGTGTCATC 1371

Qy   1501 ATCACCGGGCACCGGAAGCCCAGTTCAAGGCCCCAGGGACGGATCTTTGGGAAACTGAAA 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1372 ATCACCGGGCACCGGAAGCCCAGTTCAAGGCCCCAGGGACGGATCTTTGGGAAACTGAAA 1431

Qy   1561 GAGGAAAACTTCTTTAACCCCAAAGAAGAGAGTGAAGCTGGAAGGCCATATCAGAGTGCCC 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1432 GAGGAAAACTTCTTTAACCCCAAAGAAGAGAGTGAAGCTGGAAGGCCATATCAGAGTGCCC 1491

Qy   1621 TCTTCCACAGCTGGCCGGGTGATTGGCAAAGGTGCAAGACCGTGAACGAACTGCAGAAC 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1492 TCTTCCACAGCTGGCCGGGTGATTGGCAAAGGTGCAAGACCGTGAACGAACTGCAGAAC 1551

Qy   1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGTAAATGAGGAGTG 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1552 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGTAAATGAGGAGTG 1611

Qy   1741 ATCGTCAGAGATTATCGGGCACTCTTCTTGTGCTAGCGACGACGAGAATCAGGAGGAA 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1612 ATCGTCAGAGATTATCGGGCACTCTTCTTGTGCTAGCGACGACGAGAATCAGGAGGAA 1671

Qy   1801 ATTGTACAACAGGTGAAGCAGCAGGAGCAGGAGAAATACCCTCAGGGAGTCGCCTCCACAGCGC 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1672 ATTGTACAACAGGTGAAGCAGCAGGAGCAGGAGAAATACCCTCAGGGAGTCGCCTCCACAGCGC 1731

Qy   1861 AGCAAGTGAGGCTCCAGCACCAGCCAAACGCAGCCAGCCAAACCAAAGACCATGTAGCCCTTCCAACAC 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1732 AGCAAGTGAGGCTCCAGCACCAGCCAAACGCAGCCAGCCAAACCAAAGACCATGTAGCCCTTCCAACAC 1791

Qy   1921 CTGACAGAATGAGACACAGGCCAAATCGGAGCAAACCAAAGACCATCTGAG 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1792 CTGACAGAATGAGACACAGGCCAAATCGGAGCAAACCAAAGACCATCTGAG 1851

Qy   1981 GAATGAGAAGTCTGCGAGGACTCTGCCGAGGCCCTGAGAACCCGAGGGGC 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1852 GAATGAGAAGTCTGCGAGGACTCTGCCGAGGCCCTGAGAACCCGAGGGGC 1911

Qy   2041 CGGAGGAGGCGGGGAAGGTCAGCGGAAGGTCAGCGGTTGCCAGAACCACCACGGCCCCGCC 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1912 CGGAGGAGGCGGGGAAGGTCAGCGGAAGGTCAGCGGTTGCCAGAACCACCACGGCCCCGCC 1971

Qy   2101 CCCCAGGGCTTCTGCAGGCTTCAGCCATCCACCATCCACTTCACTCTCCTGAA 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1972 CCCCAGGGCTTCTGCAGGCTTCAGCCATCCACCATCCACTTCACTCTCCTGAA 2031

Qy   2161 CTCCCACGACGCTATCCCTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGC 2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2032 CTCCCACGACGCTATCCCTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGC 2091

Qy   2221 AAAATGCACACCCTTTTCTGTGGCAAATGTCTTCTGTGTACATATTAGAAA 2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2092 AAAATGCACACCCTTTTCTGTGGCAAATGTCTTCTGTGTACATATTAGAAA 2151

Qy   2281 GGGAAGGATGTTAAGATATGTAACTCCAAATTCCAATTTTAATCAATTATTAATTAT 2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2152 GGGAAGGATGTTAAGATATGTAACTCCAAATTCCAATTTTAATCAATTATTAATTAT 2211

Qy   2341 TTTAGAAATAATAATATAT CAAATAACTCAACTAACTCCAATTTTTAATCAATTATTAATTTT 2400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2212 TTTAGAAATAATAATATAT CAAATAACTCAACTAACTCCAATTTTTAATCAATTATTAATTTT 2271
```

```
Qy   2401 TTTTTCTTTTTAAAGAGAAAAGCAGGCTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG 2460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2272 TTTTTCTTTTTAAAGAGAAAAGCAGGCTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG 2331

Qy   2461 GTCTCACGCGTGTGAGAGAGCAGGAGAGAGAGCGAGAGCTTTGAGGCGAGAGCCACCGGACACAAATTCACCCAGAGGGAAAT 2520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2332 GTCTCACGCGTGTGAGAGAGCAGGAGAGAGAGCGAGAGCTTTGAGGCGAGAGCCACCGGACACAAATTCACCCAGAGGGAAAT 2391

Qy   2521 CTGGTCGGAAGGACACTCACGGCAGTTCGGATCACCTGTGTATGTCAACGAGAGGGATA 2580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2392 CTGGTCGGAAGGACACTCACGGCAGTTCGGATCACCTGTGTATGTCAACGAGAGGGATA 2451

Qy   2581 CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTCATGCCTCGTCTAGCTCACTACACCCATT 2640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2452 CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTCATGCCTCGTCTAGCTCACTACACCCATT 2511

Qy   2641 TCTCTTGCTTCACAGGTTTTAAACTGTTTTTGCATACTGTTTATATAATTCTCTGTCT 2700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2512 TCTCTTGCTTCACAGGTTTTAAACTGTTTTTGCATACTGTTTATATAATTCTCTGTCT 2571

Qy   2701 CTCTCTGTTTATCTCTCCCTCCCCTCCCCTCCCCTCCCCTCTTCTTCTCCATCTCCATCTCCATTCTTTTGAA 2760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2572 CTCTCTGTTTATCTCTCCCTCCCCTCCCCTCCCCTCCCCTCTTCTTCTCCATCTCCATCTCCATTCTTTTGAA 2631

Qy   2761 TTTCCTCATCCTCCAATCCCGTATCCACCGGTATCTCACGCACCCCCCCCCCCCCAGGCAAAGCA 2820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2632 TTTCCTCATCCTCCAATCCCGTATCCACCGGTATCTCACGCACCCCCCCCCCCCCAGGCAAAGCA 2691

Qy   2821 GTGCTCTGAGTATCACATCACACACAAAAGCGAAACACACAACACACCAGCCTCAA 2880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2692 GTGCTCTGAGTATCACATCACACACAAAAGCGAAACACACAACACACCAGCCTCAA 2751

Qy   2881 CTTACACTTGGTTACTCAAAGAACAAGAGTCAATGGTCACTTGTCTCTAGCGTTTTGGAAG 2940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2752 CTTACACTTGGTTACTCAAAGAACAAGAGTCAATGGTCACTTGTCTCTAGCGTTTTGGAAG 2811

Qy   2941 AGGAAAACAGGAACCCACCAAACCCACCAACCAATCAACCAACCAACAAAGAAAAATTCCAACAATG 3000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2812 AGGAAAACAGGAACCCACCAAACCCACCAACCAATCAACCAACCAACAAAGAAAAATTCCAACAATG 2871

Qy   3001 AAAGAATGTATTTTGTCTTTTGCATTTTGCATTTTGCTATAAGCCATCAATATTCAGCAAAAATG 3060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2872 AAAGAATGTATTTTGTCTTTTGCATTTTGCATTTTGCTATAAGCCATCAATATTCAGCAAAAATG 2931

Qy   3061 ATTCCTTTCTTTAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCC 3120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2932 ATTCCTTTCTTTAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCC 2991

Qy   3121 CAGGGCGTTAAATTCACAGATTTTTTTAACGAGAAAAACACACAGAAGAAGCTACCTCAG 3180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2992 CAGGGCGTTAAATTCACAGATTTTTTTAACGAGAAAAACACACAGAAGAAGCTACCTCAG 3051

Qy   3181 GTGTTTTTACCTCAGCACCTTGCTCTTGTGTTCCCTTAGAGATTTTGTTAAAGCTGATAG 3240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3052 GTGTTTTTACCTCAGCACCTTGCTCTTGTGTTCCCTTAGAGATTTTGTTAAAGCTGATAG 3111

Qy   3241 TTGGAGCATTTTTTATTTTTTAATAAAAAATAAGATATCACT 3300
           ||||||||||||||||||||||||||||||||||||||||||
Db   3112 TTGGAGCATTTTTTATTTTTTAATAAAAAATAAGATATCACT 3171

Qy   3301 GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTGTTCTTCCGCT 3360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3172 GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTGTTCTTCCGCT 3231

Qy   3361 AGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3412
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3232 AGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3283
```

RESULT 4
US-09-899-651-8
; Sequence 8, Application US/09899651

```
; Patent No. 6576756
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
; TITLE OF INVENTION: Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/899,651
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/09/061,709
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 8
; LENGTH: 3283
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-899-651-8

Query Match           92.1%;    Score 3143;   DB 4;   Length 3283;
Best Local Similarity 96.2%;    Pred. No. 0;
Matches 3283;  Conservative 0;  Mismatches 0;  Indels 129;  Gaps 1;
```

```
Qy     1 GGCAGCGGAGGAGGGGAGGAGGCGCCGGGTACCGGCGGGGAGCGGGGAGCCGGCTCTCGGGG  60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GGCAGCGGAGGAGGGGAGGAGGCGCCGGGTACCGGCGGGGAGCGGGGAGCCGGCTCTCGGGG  60

Qy    61 AAGAGACGGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCCGCCCCGAC 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 AAGAGACGGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCCGCCCCGAC 120

Qy   121 GACCTTCCGGCAGCTCTTTGGGACAGGAAGCTGCCCCTGCCCTGCGGGACAGGTCCTGCTGAAG 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GACCTTCCGGCAGCTCTTTGGGACAGGAAGCTGCCCCTGCCCTGCGGGACAGGTCCTGCTGAAG 180

Qy   181 TCCGGCTACGCCTTCGTGGACTACCCGGACTACCCGGGCCGCGGGGCCGGGAGCAGGAACTCAGAGACC 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TCCGGCTACGCCTTCGTGGACTACCCGGACTACCCGGGCCGCGGGGCCGGGAGCAGGAACTCAGAGACC 240

Qy   241 CTCTCGGGTAAAGTGAATTGCATGGGAAAAATGCATGGGAAGTTGAATCATGGAAGTTGATTACTTCAGTTCTAAA 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CTCTCGGGTAAAGTGAATTGCATGGGAAAAATGCATGGGAAGTTGAATCATGGAAGTTGATTACTTCAGTTCTAAA 300

Qy   301 AAGCTAAGGACGCAGGAGAAAAATTCAGATTGGGACAGTGGAACATGGAACAAGCAAAAATAGCCATG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AAGCTAAGGACGCAGGAGAAAAATTCAGATTGGGACAGTGGAACATGGAACAAGCAAAAATAGCCATG 360

Qy   361 TTGGATGGACTTTGCTCAATATGGGAACACGTCACATATGGAACAAGAGAAGAAAATAGCCATG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TTGGATGGACTTTGCTCAATATGGGAACACGTCACATATGGAACAAGAGAAGAAAATAGCCATG 420

Qy   421 ACAGAAACCGCCGTTGTCAACGTGCAACAAGAGAAGAAAATAGCCATG 480
           |||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ACAGAAACCGCCGTTGTCAACGTCACATATGGAACAAGAGAAGAAAATAGCCATG 480

Qy   481 GAGAAGGTAAGCGGGCATCCAGTTTGAACGGTGAGCTCCCCTTCCGCCCCTCAGCGTGGGGACCACTCTTCCCGGAT 540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GAGAAGGTGAGCGGGCATCCAGTTTGAACGGTGAGCTCCCCTTCCGCCCCTCAGCGTGGGGACCACTCTTCCCGGAT 540

Qy   541 GAAGAGGTGAGCTCCCCTTCGCCCCTCGCCCCCAGCGTGGGGACCACTCTTCCCGG 600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GAAGAGGTGAGCTCCCCTTCGCCCCTCGCCCCCAGCGTGGGGACCACTCTTCCCGG 600

Qy   601 GAGCAAGGCCACGCCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCCG 660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GAGCAAGGCCACGCCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCCG 660

Qy   661 ATCCTGGTCCCCACCCAGTTTGTTGGTGGTGCCATCATCGGAAAGGAGAGGGGCTTGACCATAAAG 720
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 ATCCTGGTCCCCACCCAGTTTGTTGGTGGTGCCATCATCGGAAAGGAGAGGGGCTTGACCATAAAG 720

Qy   721 AACATCACTAAGCAGCAGACCAGACCCAGTCCGGGTAGATATCCATAGAAAGAGAACTCTGGAGCT 780
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 AACATCACTAAGCAGCAGACCAGACCCAGTCCGGGTAGATATCCATAGAAAGAGAACTCTGGAGCT 780

Qy   781 GCAGAGAAGCCTGTCACCATGCCACCCCAGAGGGACTTCTGAAGCATGCCGCATG 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GCAGAGAAGCCTGTCACCATGCCACCCCAGAGGGACTTCTGAAGCATGCCGCATG 840

Qy   841 ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAGACCAAACTAGCCGAAGAGATTCCTCTG 900
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAGACCAAACTAGCCGAAGAGATTCCTCTG 900

Qy   901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGGCAGAAATTTG 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGGCAGAAATTTG 960

Qy   961 AAGAAAATTGAACATGAAACACAGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 AAGAAAATTGAACATGAAACACAGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020

Qy  1021 ATATACAACCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTTGTGCCAGTGCT 1080
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 ATATACAACCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTTGTGCCAGTGCT 1080

Qy  1081 GAGATAGAGATTATGAGAAGTCGCGTGAGAGTGCTTTGAAAAATGTATATGCTGGCTGTTAAC 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 GAGATAGAGATTATGAGAAGTCGCGTGAGAGTGCTTTGAAAAATGTATATGCTGGCTGTTAAC 1140

Qy  1141 CAACAAGCCAATCTGATCCGCACTTCAGCGTTGAACCTCAGCGCACTCAGGTTGAACCTCAGCGCACTCAGGTTGAACCTCAGCGCACTCAGGTTCAACAGGA 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 -------------------------------------------------------------- 1140

Qy  1201 CTGTCGGTGCTATCTCCACCAGCAGCAGGGGCCCCGCGGAGCTCCCCCGCTGCCCCCTACCAC 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 -------------------------------------------------------------- 1140

Qy  1261 CCCTTCACTACCACCTCCGGATACTTCTCCAGCCTGCTGACCCGTACCCCCATCACCAGCC 1320
           ||||||||||||||||||||||||||||||||||||||    ||||||||||||||||||
Db  1141 ---------ACCACTCCGATACTTCTCCAGCCTGCTGACCCCGTACCCCCATCACCAGCC 1191

Qy  1321 TTCCCGCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTTTCATCCCAACCCAG 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1192 TTCCCGCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTTTCATCCCAACCCAG 1251

Qy  1381 GCTGTGGGCGCCATCATCGGGAAGAAGGGGCACACATCAAACAGCTCGGCGAGATTCGCC 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1252 GCTGTGGGCGCCATCATCGGGAAGAAGGGGCACACATCAAACAGCTCGGCGAGATTCGCC 1311

Qy  1441 GGAGCCTTATCAGATTGCCCCGGAAGCCCAGTTCAAGGCCCAGGGGACGATGGTCATC 1500
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1312 GGAGCCTTATCAGATTGCCCCGGAAGCCCAGTTCAAGGCCCAGGGGACGATGGTCATC 1371

Qy  1501 ATCACCGGGCCACCGGAAGCCAGTTCAAGGCCCAGGGGATCTTTGGGAAACTGAAA 1560
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1372 ATCACCGGGCCACCGGAAGCCAGTTCAAGGCCCAGGGGATCTTTGGGAAACTGAAA 1431

Qy  1561 GAGGAAACTTCTTTAACCCCAAAGAAGATTCCTACATCCCGGAT 1620
           |||||||||||||||||||||||||||||||||||||||||||
Db  1432 GAGGAAACTTCTTTAACCCCAAAGAAGATTCCTACATCCCGGAT 1491

Qy  1621 TCTTCCACAGCTGCTGGGTGATTGGCAAGGAGACCGTGAACGAACTGCAGAAC 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1492 TCTTCCACAGCTGCTGGGTGATTGGCAAGGAGACCGTGAACGAACTGCAGAAC 1551

Qy  1681 TTAACCAGTGCGAAGTCATCGTCCTCGTGACCAAACGCCAGAGTGAAAATGAGGAAGTG 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1552 TTAACCAGTGCAGAAGTCATCGTCCTCGTGACCAAACCCCAGAGTGAAAATGAGGAAGTG 1611

Qy  1741 ATCGTCAGAATTATCGGGACACTTCTTTGCTAGCCAGACTGACAGCGCAAGATCCAGGGAA 1800
```

(DNA sequence alignment — SEQ ID comparison, Db/Qy rows with match bars; positions 1612–3412)

RESULT 5

US-09-643-597-347

; Sequence 347, Application US/09643597

; Patent No. 6426072

; GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C11
; CURRENT APPLICATION NUMBER: US/09/643,597
; CURRENT FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 369
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 347
; LENGTH: 1740
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-643-597-347

Query Match 20.9%; Score 712; DB 4; Length 1740;
Best Local Similarity 64.5%; Pred. No. 5.2e-189;
Matches 1156; Conservative 0; Mismatches 555; Indels 81; Gaps 3;

```
RESULT 6
US-09-542-615A-347
; Sequence 347, Application US/09542615A
; Patent No. 6518256
; GENERAL INFORMATION:
;   APPLICANT:  Wang, Tongtong
;   APPLICANT:  Fan, Liqun
;   APPLICANT:  Kalos, Michael D.
;   APPLICANT:  Bargur, Chaitanya S.
;   APPLICANT:  Hosken, Nancy A.
;   APPLICANT:  Fanger, Gary R.
;   TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THERAPY
;   TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;   FILE REFERENCE:  210121.455C8
;   CURRENT APPLICATION NUMBER:  US/09/542,615A
;   CURRENT FILING DATE:  2000-04-14
;   NUMBER OF SEQ ID NOS:  350
;   SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 347
;   LENGTH: 1740
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-542-615A-347

Query Match     20.9%;  Score 712;  DB 4;  Length 1740;
Best Local Similarity  64.5%;  Pred. No. 5.2e-189;
```

Matches 1156; Conservative 0; Mismatches 555; Indels 81; Gaps 3;

(nucleotide sequence alignment, Qy = query, Db = database; positions 1–1855)

```
Qy   73 ATGAACAAGCTTACATGGGAACCTGGGAACCTGAGCCGCCGGAACCTCAGCGACCTCGGCAG 132
           |||||| |||||||||  ||||| ||| || |||||||||||| |||| |||||||
Db    1 ATGAACAAACTGTATATGGAAAACCTCAGCGAGAGACGCGGCCCCTCGGACCTAGAAAGT  60

Qy  133 CTCTTTGGGACAGGAGAGCTGCCCCTGGCCGGGACAGGTCTGCTGAAGTCCGGCTACCGC 192
           |||| ||||||||||   ||||| ||||  ||||||||| ||||||||| ||||||||||
Db   61 ATCTTCAAGGACGCGCCAAGATCCCGGTGTGTCGGGACTGGCTGGAGACTGGCTACCGCG 120

Qy  193 TTCGTGGACTACCCCGACCAGAACTGGGCCATCGGCCATCGAGACCCTCTCGGGTAAA 252
            ||||||| |||||||| ||||||||| |||||||  ||||||  ||||||||||||||||
Db  121 TTCGTGGACTGCCCGGACGAGAGCCCTCAAGGCCCATCAAGGCCGCTTTCAGGTAA 180

Qy  253 GTGGAATTGCATGGGAAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGTAAGGAGC 312
           |||||| ||  | ||||| |||| |||  || | |||| ||||  ||| |||| ||| |||
Db  181 ATAGAACTGACGGGAAAAACCCATAGAAGTTGAGCACTCGGTCCCAAAAAAGGCAAAGAT 240

Qy  313 AGGAAAATTCAGATTCGAAACATCCCTTCCACCTGCAGTGGGAGGTGTTGGATGGACTT 372
           || |||||||| || |||||| ||||| |   |||| ||||||||||||||||||
Db  241 CGGAAACTTCAGATACGAAATATCCCGCCCTCATTTACAGTGGGAGGTGCTGGATAGTTTA 300

Qy  373 TTGGCTCAATATGGGACAGTGGAGAATGTCAAACAAGTCAACACCAGACACAGAAACCGCC 432
           ||||||||||||  |   |||| ||||| ||||||||| || |||||  |||||||||
Db  301 CTAGTCCAGTATGGAGTGTGAGAGTGTCTGAGCAAGTGAACACTGACCTCGGAAACTGCA 360

Qy  433 GTTGTCAAGCTCACATATGACAAGAGAGAAGCAAAATAGCCATGAGAAGCTAAGC 492
           |||| |||| |||||  |||   |||  ||| ||| ||||| |||||||  ||| ||
Db  361 GTTGTAAATGTAACCTATTCCAGTAGGAAGGACCAAGCTCAGGCTCAAGGCTCAGG 420

Qy  493 GGGCATCAGTTTGAGAACTACTCCTTCAAGATTTTCCTACATCCCGGATGAAGAGGTGAGC 552
           |||||||||||  |||| |||||  |||||||  ||||  ||||| ||||||||||||||
Db  421 GGATTTCAGTTGAGAGAATTTCACCTTGAAAGTAGCCTATAATCCCTGAATGAAACGCGCCC 480

Qy  553 TCCCCTTGCCCTCAGCGAGCCCAGCGAGCCCGGGGACCACTGTTCCCGGGAGCAAGGC--- 609
           ||||  | ||||| ||| || |||  |    |||||||| |  ||||   ||||||
Db  481 CAGCAAAACCCCTTGCAGCAGCCCGGAGTCGCGCCGGGCTTGGGCAGAGGGGGCTCCTCA 540

Qy  610 ------CACGGCCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTCCCGCTGCGGATC 663
                 || |||||||||||||  ||||| ||||| ||||| |||||||  ||||||||||
Db  541 AGGCAGGGGCTCTCCAGGATCCGTATCCAAGCAGAACAAACCATGTGATTTGCCTCTGCGGCCTG 600

Qy  664 CTGGTTCCCACCCAGTTTGTTGGTGCCATCATCATCGGAAAGGACTTGACCATAAAGAAC 723
           |||||| |||||  |||||||| |||||||||| |||| ||||| | |||  | ||||||
Db  601 CTGGTTCCCACCCAATTTGTTGTTGGCCATCATCATAGGAAAGAGAAGTGCCACCATTCGGAAC 660

Qy  724 ATCACTAAGACGACCCAGTCCCGGTAGATATCCAGAAAGAGAGAACTCTGAGCTGCA 783
           || ||| |||||| ||||| | || || ||||| |||||||| |||||||||||||
Db  661 ATCACCAAAGACCACCCAGTCTAAAATCGATGTCCACCGTAAGAAAATGCGGGGGTGCT 720

Qy  784 GAGAAGCCTGTCACCATCCATGCCACCCAGAGAGGGACTTCTGAAGCATGCCGCATGATT 843
           ||||| ||| ||  | ||||| ||| || || ||||||| ||| | | ||| |||| |
Db  721 GAGAAGTCGATTACTATCCTCTACTCCTGAGGCCACCTCTGCGGCTTGTGTAAGTCTATT 780

Qy  844 CTTGAAATCATGCAGAAGAGAGGCAGATGAGACCCAAACTAGCCAGAGATTCCTCTGAAA 903
           ||||| || | ||| ||||||||| ||| | ||| ||  |||||||||||||||  | |||
Db  781 CTGGAGATTATGCATAGGAAGCTCAAGATATAAAATTCACAGAGAGATCCCCTTGAAG 840

Qy  904 ATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTGAAG 963
           || |  | | ||||  || ||||||||||||| |||||||||||| || ||| |||||||
Db  841 ATTTTAGTCATAATAACTGGTTGGTTGGAAGAGGTGTAAAAGAAGAAGAAATCTTAAA 900

Qy  964 AAAATTGACACATGAAAACAGGAGCCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATA 1023
           |||||| || | ||| | || |||  ||| ||  |||||||| |  ||||| |||  |||
Db  901 AAAATTGAGCAGCAAGACACAGCACACTAAAATCACCGATATCTCCATTGCAGGAATTGACCCTG 960

Qy 1024 TACAACCCGGAAAGAACCATCACTGTGAAGGGACAGTTGAGGCCTGTGCCAGTGCTGAG 1083
           ||||||| | ||| |||| |||| || ||| |||| | || ||||  |||| || |||
Db  961 TATAATCCAGAACGCACTATTACAGTTAAAGGCAATGTGGAGACAATGTGTCCAAAGTCGAG 1020

Qy 1084 ATAGAGATTATGAAGCAGAAATCAGGAGTCTTATGAAAATGATAATGCTGGCTGTTAACCAA 1143
           |||| ||| |||| ||||| |||| |  | |||| ||||| ||||| | ||  ||| |||
Db 1021 GAGGAGATCATGAAGAAAAGAATAACAGGAGT CTTATGAAAATGATATTTCTCATGAATCTT 1080
```

```
Qy 1144 CAAGCCAATCTGATCCGACGGGTTGAACCTCAGCGACGCGGCCATCTTTCAACAGGACTG 1203
           |||| |  | || ||||  | || |  |  |||||| | |  |||| ||||||||||||||
Db 1081 CAAGCACATTTAATTCCTGGATTAAATCTGACGCCGCCTTGGGTCTCGTTCCCACCCACTTCA 1140

Qy 1204 TCCGTGTCATCTCCACCAGCAGCAGCAGCCGCGGAGCTCCCCCGCTCCACCACCCC----- 1263
           |||||| |||||||| ||||||   |||  || |  ||
Db 1141 GGGATGCCACCTCCCACCTCAGGGCCCCCCTTCAGCCATGACTCCT---------------  1185

Qy 1264 TTCACTACCACCTCCGGATACTTCTCCAGCCTGTCACCCCATCACCAGTTTGGCCCGTTC 1323
                                            ||||| |||| ||   | ||||||||||
Db 1186 ------------------------------CCCTACCCGCAGTTTGAGC----------  1204

Qy 1324 CCGGCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTTCATCTGTTTATCCCAGCTCTCATCA 1383
           |||                ||||||| ||| ||||  |||  ||||| ||||| ||||||||
Db 1205 -----AATCAGAAACGGAGACTGTTCATCTGTTCATCCTGTTCATCTGTTATTTATCCCAGCTCTCATCA 1248

Qy 1384 GTGGGGCCATCATCGGAAGAAGAGGGGCACACATCAAAACAGCTGGCGGAGATTCGCCCGGA 1443
           ||| ||| ||||| | |||||||||  | |||| |||| ||| |||||||||||  ||||||
Db 1249 GTCGGTGCCATCATCGGCAAGCGAGGGCCAGCATCAAGAGCTTGAAGCTCAGAGTGCCATCC 1308

Qy 1444 GCCTCTATCAAGATTGCCCCTCGCGAAGGCCCAGAGCTCAGCGAAGGATGGTCATCATC 1503
           ||||||||||||  |||| || | ||||| |||||| | ||| |||  | || ||  ||
Db 1309 GCTTCATTAAGATTGCTCCAGGAAGCCCAGATGCTAAAGTGAGGATGGATGGTGATTATC 1368

Qy 1504 ACCGGGCCACCGGAAGCCCAGTTCAAGGCCCCAGGAGACGGATCTTTGGGAACTGAAAGAG 1563
           || ||||| ||  |||  ||||||||||  ||| || |||| |||  |||  ||| ||| |
Db 1369 ACTGGACCACCAGGAGCTCAGTTCAAGGCTCAGGAGAATTTATGGAAGAAATTAAAGAA 1428

Qy 1564 GAAAACTTCTTTAACCCAAAGAAGAGTGAAGCTGGAAGCGCATATCAGAGTGCCCTCT 1623
           |||||||| ||| ||||| |||||||| | |||  |  | || | ||| ||| ||||
Db 1429 GAAAACTTTGTTAGTCCTAAAGAAGAAGGTGAAACTTGAAGCTCATATCAGAGTGCCATCC 1488

Qy 1624 TCCACAGCTGGCGCGGGTGATTGGCAAGGTGCAAGAGTGTGACCGTGAGCAGAACTTA 1683
           |||| |  | |  | || |||| ||  ||||| || || | ||  | ||  ||| |||
Db 1489 TTTGCTGCTGGAGAGTTATTGGAAAAGGAGCAAAACGGTGAATGAACTTCAGAATTTG 1548

Qy 1684 ACCAGTCAGAAGTCATCGTGTCCTCGTCGTCGGATCCCGCTCGCGGATC 1743
           ||  || ||||| | | ||||  |  |||||  |  |||||||||||||||
Db 1549 TCAAGTGCAGAAGTTGTGTTCCCTGTGTCCTCGTGACGTCACCAGTGGTT 1608

Qy 1744 GTCAGAATTATCGGGCACTTCTTTTGCTGGCACTCTTGTCTTCTTGTGTCTTCAGGAAATT 1803
           ||||| |||| |   || | ||||||  | ||| | ||  ||| || |||||||||||||
Db 1609 GTCAAAATAACTGGTCACTTCTATGTGGCCAGCAGTTGCCCAGAGAGAAAAATTCAGGAAATT 1668

Qy 1804 GTACAACAGTGAAGCAGCAGGAGCAGAAATAACCCTCAGGGAGTCGCCTCAC 1855
           ||  ||  ||||| |  | | ||| || ||| || |  |||||||| |||| 
Db 1669 CTGACTCAGGTAAGCTGGTAAAGTCTGCAAAGTCGACCAC 1720
```

```
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-606-421B-347

Query Match      20.9%;  Score 712;  DB 4;  Length 1740;
Best Local Similarity  64.5%;  Pred. No. 5.2e-189;
Matches 1156;  Conservative  0;  Mismatches 555;  Indels  81;  Gaps  3;

Qy     73 ATGAACAAGCTTTACATGGGACCTAGAGCCCGCCGTCACCGGCGACGACTTCCGGCAG 132
          |||||||||||||||||| ||||| |||  || ||||||||| ||||||||||||||
Db      1 ATGAACAAACTGTATATCGGAAACATCAGCGAGAACGCGCCCCTTCGGACCTAGAAAGT 60

Qy    133 CTCTTTGGGGACAGGAAGTGCCCTGGCGGGACAGGTCTGCTGAAGTCCGGCTACGCC 192
          |||| ||||| |||||||||  | |||| ||||||||| |||||||||||||||||||
Db     61 ATCTTCAAGGACGACGACGATCCCGGTGTCGGGACAGGTCTGCTGAAGTCCGGCTACGCG 120

Qy    193 TTCGTGGACTACCCCGACCAGGAACTGGGCCATCCGGCCATCGAGACCCTCTCGGGTAAA 252
          |||||||||||||||||||||||| |||||||| ||||| |||||||| |||||||||
Db    121 TTCGTGGACTACCCCGACCAGGAGCTGGGCCTTCAAGGCCATCGAGACCGCTTTCAGGTAAA 180

Qy    253 GTGGAATTGCATGGGGAAAATCATGGAAGTTGATTACTGGATCTCTAAAAAGTAAGGAGC 312
          |||||||||||| | |||||||  ||||||||||||||| | |||| |||||||||||
Db    181 ATAGAACTGCACGGGGAAACCCCATAGAAGTTGAGCACTCGTCCCAAAAAGGCAAAGGATT 240

Qy    313 AGGAAAAATTCAGATTGAAACATCCTCCTCACGTGGGAGGTGTTGGATGGACTT 372
          |||||||| |||||| |||| ||||||| |||||||||  ||||  || |||||
Db    241 CGGAAACTTCAGATACGATGAAATATCCCGGCCTCATTTACAGTGGGAGGTGCTGGATAGTTTTA 300

Qy    373 TTGGCTCAATATGGGACAGTGGAGAATGTGGACAAGTCAACACAGACACCGCC 432
          ||||||||||  |||||| | |||||||||||| || |||| ||||||||| ||
Db    301 CTAGTTCCAGTATGGTGGTGGGAGAGCTGTGAGCAAGTGAACACTCGGAAAACTGCA 360

Qy    433 GTTGTCAAGTCACATATGGCAACAGAAGAAGCAAAAATAGCCATGATGAGGAAGCTAAGC 492
          ||| |||||| |||||||| ||||||| ||||||| ||||||||||| |||| || |||
Db    361 GTTGTAAATGTAACCTATTCCAGTAAGGACCAAGCTAGAAGCAAGAACTGAAT 420

Qy    493 GGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCACATCCCGATGAGAGAGGTGAGC 552
          ||| |||||||||||||||| |||||||||||  ||| |||||||| |||||||||||
Db    421 GGATTTCAGTGAGTTAGAGAATTTCACCTGAAAGTAGCCTATATCCTGATGAAACGCGCCGCC 480

Qy    553 TCCCCTTCGCCCCCTCAGCGAGCCCAGCGTGGGAACCACTCTTCCCGGAGCAAGGC 609
          |||| | |||||||||||||||||||| ||||| |||||| |||| ||||| ||||
Db    481 CAGCAAAACCCCTTGCGCAGCGCCCGAGTGCCCGAGTTGGGCAGAGGGGCTCCTCA 540

Qy    610 ------CACGGCCCCTGGGGGCCACTTCTCAGGCCGACAGATTGATTTCCCGCTGCGGATC 663
                ||||||| ||| ||| |||||||||||||| |||||| |||| ||||||||||||
Db    541 AGGCAGGGGTCTCCAGGATCCGTATCCGTATCCAAGCGAAACCATGTGATTTGCCTCTGCCGCCTG 600

Qy    664 CTGGTCCCCACCCAGTTGTTGGAGCCCAGCGTTCGACCACTCTTGACCCATAAAGAAC 723
          ||||| |||||||||||||| ||||||||| | ||||||| |||||||| | ||||||
Db    601 CTGGTTCCCCAATTGTTTGGAGCCCATCATAGGAGCCCATCATAGGAAAAGAAGG 660

Qy    724 ATCACTAAGGCAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAAGCAACTCTGGAGCTGCA 783
          ||||| || | |||||||| | |||||| ||||| ||||||||||| | |||||| || |||
Db    661 ATCACCAACAGACCCAGTCTAAAATCGATTGTCCACCGTAAAGAAAATTGCGGGGGCTGCT 720

Qy    784 GAGAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATGGATT 843
          |||||| || |||||| |||||| || | ||||||||||| |||||| || ||||||||
Db    721 GAGAAGTCGATTACTACTATCCTCCTACTCCTGAAGGCACCTCTGAAGCCTTGTAAGTCTATT 780

Qy    844 CTTGAAATCATGCAGAAAGAGGCCAGATGAGAACCAAAACTAGCCGAAGAGATTCCTCTGAAA 903
          |||| || |||||| |||||| |||| |||||| | || |||||||| |||||||| ||||
Db    781 CTGGAGATTATGCATAAGGAGCTCAAGATAATAAAAATCACAGAAGAGATCCCCTTGAAG 840

Qy    904 ATCTTGGCCACACAATGCCTGGTTCGGAAGACTGATTGGAAGAGCAGAAAATTTGAAG 963
          | ||| |||  ||| | || ||||||| |||||||||   |||||||||||| |||||
Db    841 ATTTTAGCTCATAATAACTTGTTGGACGTCTTATTGGTAAAGAAGAAGAAGAAATCTTAAA 900

Qy    964 AAAATTGAACATGAAACATGAAAGATAACAATCCATCTCATCTTTGCAGGATTTGAGCATA 1023
          |||||| |||| ||| |||||| |||||| || ||||| |||| |||||| |||||||||
Db    901 AAAATTGAGCAGAAGCAACAGACAACTAAAATCACGATATCTCCATTGCAGGAATTGACGCGCTG 960

Qy   1024 TACAACCCGGAAGAAGACCATCACTGTGAAGGGCACAGTGAGGCCTGTGCCAGTGCTGAG 1083
          || |||||||||||||||| | || |||||||  ||||||||||| || ||||| |||||
Db    961 TATAATCCAGAACGCACTATTACAGTTAAAGGCAATGTTGAAAATGATGATGCTGGCTGAG 1020

Qy   1084 ATAGAGATTATGAGAGAGCTGCGTGCGAGCGAGCTGCCTTTGAAAAATGATGATGCTGGCTGAG 1143
          |||||||||| |||||| ||| |||| |||||||| ||||||||||||||||||||||||||
Db   1021 GAGGAGATCATGAAGAAAATACAGGGAGTCTTATGAAAATGATTATTGCTTCATGAATCTT 1080

Qy   1144 CAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTCAACAGAGACTG 1203
          |||| ||||| |||||||||| ||||| | || | || ||||| || |||||||||||||
Db   1081 CAAGCACATTTAATTCCTGGATTAAATCTGAGCGCCTTGGGTCGTTCCCCCACTTCA 1140

Qy   1204 TCCGTGTATCTCCACCAGCAGGAGCGGGCTGTGAAGTCCGGCTGCCCCCGGTACCACCCC 1263
          |||| ||| ||||||||||||||||||||  | || |||  ||||  ||| |||||||||
Db   1141 GGGATGCCACCTCCCACCTCCAGGGCCCTTCAGCCATGACTCCT------ 1185

Qy   1264 TTCACTACCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTGCCCCGGTTC 1323
          |||||||||||||||||||||| |||||||| |||||||| |||||||| |||||||
Db   1186 ------------------------CCCTACCCGCAGTTTGAGC------ 1204

Qy   1324 CCGCATCATCACTCTTATCCAGAGCAGGAGGATTGTGAATCCTTCATCCCACCCAGGCT 1383
          ||||||||||||||||||||||||||||||||||||  ||| || || |||||||||||
Db   1205 ------------AATCAGAAACGAGACTGTTCATCTGTTATCCCAGCTCTATCA 1248

Qy   1384 GTCGGGCGCCATCATCGGGAAGAAGAGGGGCCACACATCAAACAGCTGGCGAGATTCGCGGA 1443
          |||| |||||||||||| ||  ||| |||||||| ||| |||||||||||| | |||||||
Db   1249 GTCGGTGCCATCATCGGCAAGCAGCGGGCCATCAAGCACTTTCTCGCTTTGCTTGGA 1308

Qy   1444 GCCTCTATCAAGATTGCCCCTGCGGAAGGCCGGAAGACGTCAGCGGAAGGATGTGTCATCATC 1503
          ||||| ||||| |||||  ||| |||||||||||||| ||||| |||||||| ||| ||||||
Db   1309 GCTTCAATTAAGATTGCTTCCAGCGGAAGCACCAGCGATGCTAAAGTGAGGATGGTGATTATC 1368

Qy   1504 ACCGGGCACCGGAAGCCCAGTTCAAGCCCAGTGACGGATCTTTGGGAAACTGAAAGAG 1563
          ||||| |||||||||  || ||||| | |||| ||| |||| ||| ||||| ||||||
Db   1369 ACTGGACCACCAGAGGCTCAGTTCAAGGCTCAGGGAAGAATTTATGGAAAAATTAAAGAA 1428

Qy   1564 GAAAACTTCTTTAACCCCAAAGAAGAGGTGAAGCTGGAAGCGCATATCAGAGTGCCCTCT 1623
          ||||||||| || || ||||||||||| ||||| || |||| |||||||||||||  |||
Db   1429 GAAAACTTTGTTAGTCCTAAAGAAGAGTGTAAGCTTCATATCAGAGTGAAATCTGCCATCC 1488

Qy   1624 TCCACAGTCGGCGGTGATTGGCAAGGTGGCAAAGGTGCGAAGACCGTGAACGAACTGAACTTA 1683
          |||| |||||  ||| || |||||| |||||||||| | ||| || || |||||||||  ||
Db   1489 TTTGTCGTGGGCAGAGTTATTGGAAAGGAGCAAAACGGTGAATGAGCTTCAGAGAATTTG 1548

Qy   1684 ACCAGTGCAGAAGTCATCGTGCCTCGTCGTGACCAAAACGCCAGATGAGAAAATGAGGAAGTGATC 1743
          |||||| |||||||| ||| || ||| |||||||| |||  ||||| || |||||| ||
Db   1549 TCAAGTGCAGAAGTTGTTGTCCCTCGTGACCAGCACACCTGATGAATGACCAAGTGGTT 1608

Qy   1744 GTCAGAATTATCGGCACTTCTTGTACTAGCCAGACTGCACAGGCCAAGATCAGGGAAATT 1803
          |||| |||| | |||||| |||| | |||| ||| || |||| |||| || |||||||||
Db   1609 GTCAAAATAACTGGTCACTTCTCATGCTTGTGCCCAGAGAGAAAAATTCAGGAAATT 1668

Qy   1804 GTACAACAGGTGAAGCAGCAGCAGAAATACCCTCAGGGAGTCGCCTCAC 1855
          ||||| ||||||||  ||||| || ||||| ||||||||| ||||| ||
Db   1669 CTGACTCAGGTAAAGCAGTAAAGCAGCACCAACAGAGAGGCTCTGCAAAGTGGACCAC 1720
```

```
;  NUMBER OF SEQ ID NOS: 8
;  SEQ ID NO 4
;  LENGTH: 4159
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
US-09-061-709-4

Query Match      20.8%;   Score 710.4;  DB 3;  Length 4159;
Best Local Similarity  64.5%;  Pred. No. 2.7e-188;
Matches 1155;  Conservative  0;  Mismatches 556;  Indels  81;  Gaps  3;

Qy    73  ATGAACAAGCTTACATCGGAACCTGGAACCTCGAGCCCCGTCACCGCCGACGACCTCCGGCAG  132
                                                                
Db   251  ATGAACAAACTGTATATCGGAAACCTCAGCGAGCAGAGCGCCGCCCCTCGGACCTAGAAAGT  310

Qy   133  CTCTTTGGGGACACAGGAAGCTGCCCCTGGCGGACAGGTCCTGCTTGGTCCGGCTACGCGC  192
                                                                
Db   311  ATCTTCAAGGACGCCAAGATCCCCGGTGTCGGGACCCTTCCTGGGGAAGACTGCTACGCG  370

Qy   193  TTCGTGGACTACCCCGACCAGAACTGGACCATGGACCATCGAGACCCTCTCGGGTAAA  252
                                                                
Db   371  TTCGTGGACTGCCCCGGACGAGAGCGTGGGCCCTCAAGGCCATCGAGGCGCTTTCAGGTAAA  430

Qy   253  GTGGAATTGCCATGGACGGAAAATCATGCAGTCGATTACTCAGTCTCTAAAAAGTAAGGAGC  312
                                                                
Db   431  ATAGAACTGCACGGGAAACCTCAGCGGTCCCAATGAGCACTCGGTCCCAAAAGGCAAAGGATT  490

Qy   313  AGGAAAAATTCAGATTCGAAACATCCCTCCACTGCAGTGGGGAGGTGTTGGATGGACTT  372
                                                                
Db   491  CGGAAAACTTCAGATACGAAATATCCGCGCCATTAACAGTGGGAGGTGCTCGGATAGTTTA  550

Qy   373  TTGGCTCAATATGGGACAGTGGGAGAATGTGGAACAAGTCAACACAGACACAGAAACCGCC  432
                                                                
Db   551  CTAGTCCAGTATGGAGTGGTGGAGAGCTGTGGAGCCAAGTGAACACTGACTCGGAAACTGCA  610

Qy   433  GTTGTCAACGTCACATATGCAACAAGAGGAGAGCAAAAATAGCCATGGAGAAGCTAAGC  492
                                                                
Db   611  GTTGTAAATGTAACCTATTCCAGTAAGGAACCAAGCTAGACAGCAGCACTGAGACAAACTGAAT  670

Qy   493  GGGCATCAGTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGATGAAGGAGGTGAGC  552
                                                                
Db   671  GGATTTCAGTTAGAGAATTTCACCTTGAAGTAGCCTATATCCCTGATGAAATGGCCGCC  730

Qy   553  TCCCCTTGCCCCCTGCGAGCCCAGCGTGGGACCACTCCCTCTTCCCCGGAGCCAAGGC---  609
                                                                
Db   731  CAGCAAAACCCCTTGCAGCAGCCCCGAGGTCGGCCCGGGAGCGAGAGGGGCTCCTCA  790

Qy   610  -----CACGGCCCTGGGGCACTTCTCAGGCCCAGACAGATTGATTTCCCGTCGCGATC  663
                                                                
Db   791  AGGGCAGGGGTCTCCAGGATCCGTATCCAAGCAGGAAAACCATGTGATTTGCCTTCGCCGCTG  850

Qy   664  CTGTGTCCCACCCAGTTTGTTGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAAC  723
                                                                
Db   851  CTCGGTTCCCCAATTTGTTGGAGCCATCATAGGAAAGAGAAGGTGCCACCATTCCGGAAC  910

Qy   724  ATCCATAAGCAGACGACCCAGTCTCCCGGGTAGTATATCCATAGAAAAGAAGAGACTTCTGGAGCTGCA  783
                                                                
Db   911  ATCACCAAACAGACCCAGTCTAAAATGCTAAAAATGCATGTGTCACCGTAAAGAAAAATGCGGGGGCTGCT  970

Qy   784  GAGAAGCCTGTCACCATCCATGCCACCCCAGGGGGACTTCTGAAGCCATGCCGATGATT  843
                                                                
Db   971  GAGAAGTGATGATTTACTACTCCTCTCTACTCCTGAAGGCCACCTCTGCGGCTTGTAAGTCTATT  1030

Qy   844  CTTGAAATCATGCGAGAAAGAGGCAGATCAGCAGGGAGAAGATTCCTCTGAAA  903
                                                                
Db   1031 CTGGAGATTATGCATAGGAGGAGCTCAAGATAAAATTCACAGAAGATCCCCTTGAAG  1090

Qy   904  ATCTTGGCACACAATGCCTTGGTGGAAGACTGATTGGAAGAGAAAGAAGGCAGAAATTTGAAG  963
                                                                
Db   1091 ATTTTAGCTCATAATAACTTTGTTGGAGACGTCTTATTGTAAAGAAGGAAGAAATCTTAAA  1150

Qy   964  AAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATA  1023
                                                                
```

```
Db  1151  AAAATTGAGCAGCAAGACACAGACACTAAAATCACGACTATCTCCATTGCATTGGAGGAATTGACGGCTG  1210
                                                                
Qy  1024  TACAACCCGGCAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGAG  1083
                                                                
Db  1211  TATAATCCAGAGACGACATATTACAGTGAAGCAGCATTAAAGCAACATGTGCCAAAGTGAG  1270

Qy  1084  ATAGAGATTATGAAGAAGCTCGGTGAGGCCTTGAAAATGATAATGTGCTGGTGTTAACCAA  1143
                                                                
Db  1271  GAGGAGATCATGAAGAAATCAGGAGTCTTATGAAAATGATATTGCTTCTATGAATCTT  1330

Qy  1144  CAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGACTTGGCATCTTTCAACAGGACTG  1203
                                                                
Db  1331  CAAGCACATTAATTCTGATTAAATCTGAACGCCTTGGGTCTGTTCCCACCCACTTCA  1390

Qy  1204  TCCGTGTATCTCCACCAGCAGGGCCCCGGCGAGTCCCCCGCGTGCCCCTACCACCCC  1263
                                                                
Db  1391  GGGATGCCACCTCCCACCTCAGGGCCCCTTCAGCCATGACTCCT-----  1435

Qy  1264  TTCACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGCCCCGTTC  1323
                                                                
Db  1436  -----CCCTACCCGCAGTTTGAGC-----  1454

Qy  1324  CCGCATCATCACTCTTATCCAGACAGGAGGAGGAGGAGGAGGATTGTGAATCTTCTCATCCCAACCCAGGCT  1383
                                                                
Db  1455  -----AATCAGAAACGGAGACTGTTCATCAGTTTATCCCAGCTCTATCA  1498

Qy  1384  GTGGGGCCATCATCGGAAGAAGGGGCACACATCAAACAGCTCGGCGAGATTCCCCGGA  1443
                                                                
Db  1499  GTCGGTGCCATCATCGGCAGCATGTCGCAGCGACATCAAGCAGCTTCTCGCTTTGCTGGA  1558

Qy  1444  GCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGAGACGTCAGCGGAAGGATGGTCATCATC  1503
                                                                
Db  1559  GCTTCAATTAAGATTGCTCCAGCGAGCACCAGAGCTAAAGTGAGGATGGTGATTATC  1618

Qy  1504  ACCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGGACGATCTTTGGGAAACTGAAAGAG  1563
                                                                
Db  1619  ACTGGACCACCACCAGAGGCTCAGTTCAAGGCTCAGGGAAGAATTTATGGAAAAATTAAAGAA  1678

Qy  1564  GAAAACTTCTTTAACCCAAAGAAGAAGTGAAGCTGGAGCGCATATCAGAGTGCCCTCT  1623
                                                                
Db  1679  GAAAACTTTGTTAGTCCTAAAGAAGAAGTGAAACTTGAACTCATATCAGAGTGCCATCC  1738

Qy  1624  TCCACAGCTGGCCGGGTGATTGGCAAGTGGCAAGACCGTGAACGAACTGCAGAACTTA  1683
                                                                
Db  1739  TTTGCTGTGCTGGCAGAGTTATTGGAAAAGGAGCAAAACGGTGAATGAACTTCAGAATTTG  1798

Qy  1684  ACCAGTGCAGAGTCATGGTGCCTCGTGCCCTCGTGCTGACCAAAACGCCAGAATGAGGAAGTGATC  1743
                                                                
Db  1799  TCAAGTGCAGAGTTGTGTCCCTGCGTGCTGACACCTGATGAGAATGACCAAGTGGTT  1858

Qy  1744  GTCAGAATTATCGGGACTTCTTTGCTGGCAGACTGCACAGCGCAAGATCAGGGAAATT  1803
                                                                
Db  1859  GTCAAAATAACTGGTCACTTCTATGCTTGCCAGGTTGCCAGAGAAAAAATTCAGGAAATT  1918

Qy  1804  GTACAACAGGTGAAGCAGCAGGAGCAGAGAATACCCTCAGGGAGTCGCCTCAC  1855
                                                                
Db  1919  CTGACTCAGGTAAAGCGGTCTGCAAAGTCTGGACCAC  1970
```

```
RESULT 9
US-09-899-651-4
;  Sequence 4, Application US/09899651
;  Patent No. 6576756
;  GENERAL INFORMATION:
;  APPLICANT: Chen, Yao-Tseng
;  APPLICANT: Gure, Ali
;  APPLICANT: Tsang, Solam
;  APPLICANT: Stockert, Elisabeth
;  APPLICANT: Jager, Elke
;  APPLICANT: Knuth, Alexander
;  APPLICANT: Old, Lloyd J.
;  TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
```

```
;  TITLE OF INVENTION:  Associated
;  TITLE OF INVENTION:  Antigen, The Antigens Per Se, And Uses Thereof
;  FILE REFERENCE: LUD 5538
;  CURRENT APPLICATION NUMBER: US/09/899,651
;  CURRENT FILING DATE: 2001-07-06
;  PRIOR APPLICATION NUMBER: US/09/061,709
;  PRIOR FILING DATE: 1998-04-17
;  NUMBER OF SEQ ID NOS: 8
;  SEQ ID NO 4
;    LENGTH: 4159
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
US-09-899-651-4
```

```
Query Match             20.8%;     Score 710.4;  DB 4;   Length 4159;
Best Local Similarity   64.5%;     Pred. No. 2.7e-188;
Matches 1155;   Conservative    0;  Mismatches  556;   Indels   81;   Gaps    3;
```

```
Qy    73  ATGAACAAGCTTTACATGGGAACCTGAGCCCGCGTCACCGCCGACGACCTCCGGCAG  132
          ||||||||||||| ||||||||| |||||| |||||  |||  ||||   |||||
Db   251  ATGAACAAACTGTATATCGGAAACCTCAGCGGAGAACGCGCCCCCTCGACCTAGAAAGT  310

Qy   133  CTCTTTGGGGACAGGAAGTGCCCCTGGCGGACAGGTCCTGCTGAAGTCCGGCTACGCC  192
          |   ||| ||| |||| ||| |||||| ||| |||| ||||| ||||| |||| ||
Db   311  ATCTTCAAGGACGGCCAAGATCCCGGTGTCGGGACCCTTCCTGGTGAAGACTGGCTACGCG  370

Qy   193  TTCGTGGACTACCCCGACCAGAACTGGGCCATCCGCGCCAGAGAACCCTCTCGGGTAAA  252
          |||||||||| |||||| ||||||||| ||||| | || |||||| |||||| ||||||
Db   371  TTCGTGGACTGCCCCGACGAGAGCTGGGCCCTCAAGGCCATCGAGGCGCTTTCAGGTAAA  430

Qy   253  GTGGAATTGCATGGGAAAATCATGGAAGTGATTACTACAGTCTCTAAAAAGTTAAGGAGC  312
          |||||||| ||  ||||||||| || |||||| || ||| ||||||||| ||||
Db   431  ATAGAACTGCACGGGAAACCCATAGAAGTTGAGCACTCGTCCAAAAAGGCAAGGATT  490

Qy   313  AGGAAAAATTCAGATTCGAAACATCCCTCACCTGCAGTGGGGTGTTGGATGGACTT  372
          |||| || || |||| ||||| ||||  ||||||||| | |||| || ||||| |||
Db   491  CGGAAACTTCAGATACGAAATATCCCGCCTCATTACAGTGGGAGGTGCTGGATAGTTTA  550

Qy   373  TTGGCTCAATATGGGACAGTGGAGATGTGGAACAAGTCAACACAGACAACCCGCC  432
          ||||| |||| ||| |  || || |  ||| |||||| || || |||| || ||||
Db   551  CTAGTCCAGTATGGAGGTGGTGGAGCTGTGAGCAAGTGAACACTGACTCCGGAAACTGCA  610

Qy   433  GTTGTCAACGTCACATATGCAACAAGAGAAGAGAGAGAAAATAGCCATGGAAGCTAAGC  492
          ||||| ||| |||| || |  ||| | || || ||||||| ||| ||  || |||||
Db   611  GTTGTAAATGTAACCTATTCCAGTAAGGACACAAGCCTAGACACTAGACAAACTGAAT  670

Qy   493  GGGCATCAGTTTGAGACTACTCCTTCAAGATTCCTACATCCCGGATGAGAGGTGAGC  552
          |||||||||||| || ||  || || ||| || ||| ||||| | || ||| |
Db   671  GGAATTTCAGTGAGATGAGAATTCACCTTGAAACTAGAGTAGCCTATATCCCTGATGAAATGGCCGCC  730

Qy   553  TCCCCTTCGCCCCCTCAGCGAGCCCAGCGTGGGACCACTCTTCCCGGGCAAGGC---  609
          ||||| || || ||| || ||||| ||| |||||  || ||||||| || ||||
Db   731  CAGCAAAACCCCTGCAGCAGCCCGAGGTCGCCCGGAGGTCGCCCGGGGGCTCCTCA  790

Qy   610  ------CACGGGGGTCTCCAGGATCCAGTCTAAAATCAACATCCACGTAAAGAAAATCGGGGGCTGCT  663
                |||||| ||||||||||| || || ||||| || ||  ||  || |||||||||
Db   791  AGGCAGGGGTCTCCAGGATCCGTATCCAAGCAGAAAACCATGATTTGCCTCTTGCGCCTG  850

Qy   664  CTGGTCCCCACCCAGTTGTTGTGTGCCATCATCGGCCATCATAGGAAGAAGAGGGCTTGAC  723
          ||||||||||||||||| || ||| ||||||| ||||||||||| |||||||||| || ||
Db   851  CTGGTTCCCAACCCAATTTGTTGGAGCCATCATAGGAAGAAAGAAGAAGGTGCCACCATTCGGAAC  910

Qy   724  ATCACTAAGCACAGAGACCCAGTCCCGGTAGATATCCAT  783
          |||| | || ||| |||||||||||||| | ||| |
Db   911  ATCACCAAACAGACCCAGTCCAGTCTAAAATCATGTCCACGTAAAGAAAATCGGGGGCTGCT  970

Qy   784  GAGAAGCCTGTCACCATCCACCCCGACGGGGACTTCTGAGCCATGCCGCATGATT  843
          ||||| ||| ||||||||| |  ||
Db   971  GAGAAGTCGATTACTATCCTTCTACTCGCGAAGGCGCACTTGAAGTCGCCTTGTAAGTCTATT  1030

Qy   844  CTTGAAATCATGCAGAAAGAGGCAGATGAGAGCCAAACTAGCCGAAGAGATTCCTCTGAAA  903
          |||||| |||||| |||  ||| |||| |||| |  || ||||| ||| |||||||||||
```

```
Qy  1031  CTTGGAGATTATGCATAAGGAAGCTCAAGATATAAAATTCACGAAGAAGATCCCCTTGAAG  1090
          |||| || || |||| ||||| |||||||| ||||| |||||||||||  |||||||||||
Db   904  ATCTTGGCACAACAATGGCTTGGTTGGAAGACTGATTGGAAGAAGGCAGAAATTTGAAG  963

Qy  1091  ATTTTAGTCATATAACTTTGTTGGACGTCTTATTGTTGGACGTCTTATTGTAAAGAAGGAGAAATCTTAAA  1150
          |||||||| |||| |||| |||| ||||| || |||  ||||||  ||||||| || ||||
Db  1151  AAAATTGAACATGAAACATGAAACAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGAG  1210

Qy  1024  TACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGAG  1083
          |||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db  1211  TATAATCCAGACGACGCTACTATTACCAGTTAAAGGCAATGTTGAGACATGTGCCAAAGCTGAG  1270

Qy  1084  ATAGAGATTATGAAGAAGCTGCGTGAAGATGATATGCTGGCTGTGTTAACCAA  1143
          |||| | ||||||||||| |||||||||||||||| |  |||| || |||||
Db  1271  GAGGAGATCATGAAGAAAATCAGAGGAGTCTTATGAAAATGATATTGCTTCTATGAATCTT  1330

Qy  1144  CAAGCCAATCTGATCCCCAGGGTTGAACCTCAGCGCACTTGAACTCAGCGGCACTG  1203
          ||||| |||| || |  |||  || ||||  ||  ||||| || | ||| |||||
Db  1331  CAAGCACATTAATTCCTGGATTAAATCTGACAGCTGGGTCTGTTCCCACCCACTTCA  1390

Qy  1204  TCCGTGCTATCTCCACCAGCAGGCTCCCCGGAGCTCCCCCGCTAGCCCTACCACCCC  1263
          ||  ||||| ||||||||| || | || ||||||  || || |
Db  1391  GGGATGCGCACCTCCACCTCAGGGCCCCTTCAGCGCCATGACTCCT--------------  1435

Qy  1264  TTCACTACCCACTCCGGATACTTCTCCGATCCCCATCACCAGTTTGGCCCGGTC  1323
          || | || ||  ||| |||||||||||||| |||||||||| || |||| |||
Db  1436  -----------------------CCCTACCCGCAGTTTGAGC--------------  1454

Qy  1324  CCGCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTCTTCATCCCAACCCAGGCT  1383
          |||||||| |||  |||||||| || || | ||| |||  || ||||||| || |
Db  1455  --------------AATCAGAAACGAGACTGTTCATCAGTTTATCAGTTTATCCAGTCTATCA  1498

Qy  1384  GTGGCGCCATCATCGGGAACGTGCCCATCCTTGGCCAGATTCGCCGGA  1443
          |||| |||||||||| | ||  || ||||| | | || || || ||
Db  1499  GTCGGTGCCATCATCGGCAGCAGGCGCCAGCACATCAAGCAGCTTCTCGCTTTGCTGA  1558

Qy  1444  GCCTCTATCAAGATGCCCCTGCCGGATGGCTCAGCGAACGTCATCATC  1503
          || ||| |||| || ||  || || |||| || || |||| |  ||||
Db  1559  GCTTCAATTAAGATTGCTCAGCGGAAGCACCAGAGTGTAAAGTGAGGTGATTATC  1618

Qy  1504  ACCGGGCACCGGAAGCCCAGTTCAAGCCATGGAGGACGATCTTTGGGAAACTGAAAGAG  1563
          || || |||| ||||| |||| |||| ||||| || |||||| |||||||| |||||||
Db  1619  ACTGGACCACCAGGCCTCAGTTCAAGGCTCAGGGAACGAATTTTATGGAAAAATTAAAGAA  1678

Qy  1564  GAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGAAGCTGGAAGCGCATATCGAGTGCCCTCT  1623
          || |||||| ||  |||| ||||||||  ||| ||||| || |||| || |||||||||||||
Db  1679  GAAAACTTTGTTAGTCCTAAAGAAGTGAAACTTGAAGCTCATATCGAGTGCCATCC  1738

Qy  1624  TCCACAGCTGCCGGGTGATTGGCAAAGTGGCAAGACGGAACTTGAGGAAGTGATC  1683
          ||  ||| || ||| |||| ||| ||  || |  ||| ||| ||||||  || |
Db  1739  TTTGCTGCGAGAGTTATTGGAAAGAAAAGAGGAGCAAAACGAATGAATGAAGAATTTG  1798

Qy  1684  ACCAGTGCAGAAGTCATCGTGCCTCGTCGACTAGCAGAGATTCCCGGTCTTGCGGATC  1743
          |||| | |||  || || |  ||||  || || |||  || ||| || ||| || |
Db  1799  TCAAGTGCGAAGTGTGTCCTCGCTGACCAGACACCTGACCAGTGCCAGTGCCAGTGGTT  1858

Qy  1744  GTCAGAATTATCGGCACTTCTTTGACCATAAAGAAC  1803
          || |||| ||  | || ||  ||| || |  |||||
Db  1859  GTCAAAATAACTGTCACTTCTATGCTTGCCAGAGAAATTCAGGAAATT  1918

Qy  1804  GTACACAGGTGAGCAGCAGGAGTCCCTCAC  1855
          || |||||| ||| | || ||||| ||
Db  1919  CTGACTCAGTAAAGCAGCAGCAGCAACAGAAGAGGCTCTGCAAAGTGACCAC  1970
```

```
RESULT 10
US-09-643-597-175
; Sequence 175, Application US/09643597
; Patent No. 6426072
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
```

```
;   APPLICANT:  Fan, Liqun
;   APPLICANT:  Kalos, Michael D.
;   APPLICANT:  Bangur, Chaitanya S.
;   APPLICANT:  Hosken, Nancy
;   APPLICANT:  Fanger, Gary R.
;   APPLICANT:  Li, Samuel X.
;   APPLICANT:  Wang, Aijun
;   APPLICANT:  Skeiky, Yasir A.W.
;   APPLICANT:  Henderson, Robert A.
;   APPLICANT:  McNeill, Patricia D.
;   TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY
;   TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;   FILE REFERENCE:  210121.455C11
;   CURRENT APPLICATION NUMBER: US/09/643,597
;   CURRENT FILING DATE:  2000-08-21
;   NUMBER OF SEQ ID NOS:  369
;   SOFTWARE:  FastSEQ for Windows Version 3.0
;
;   SEQ ID NO 175
;   LENGTH: 4181
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3347)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (3502)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (3506)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (3520)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (3538)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (3549)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (3646)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (3940)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (3968)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (3974)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (4036)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (4062)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (4080)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (4088)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (4115)
;   OTHER INFORMATION: n=A,T,C or G
US-09-643-597-175
```

Query Match   20.8%;   Score 710.4;   DB 4;   Length 4181;

Best Local Similarity   64.5%;   Pred. No. 2.8e-188;
Matches 1155; Conservative 0; Mismatches 556; Indels 81; Gaps 3;

```
Qy       73  ATGAACAAGCTTACATCGGGAACCTGAGCCCCGCGTCACCGCACGACCTCCCGCAG  132
             |||||||||||| ||||||| ||||||||||  || |  |||   ||| |||||||
Db      251  ATGAACAAACTGTATATCGGAAACCTCAGCGAGAACGCCGCCCCTCGGACCTAGAAAGT  310

Qy      133  CTCTTTGGGACAGGAGAGTGCCCTGCCCTGGCGGACAGGTCCTGCTGAAGTCCGGCTACGCC  192
             ||| |||   |||| |  ||| || ||| ||||    ||||| |||||| |||| ||| |
Db      311  ATCTTCAAGGACGCGCAAGATCCCGGTGTCGGGACCCCTTCCTGGTGAAGACTGGCTACGCG  370

Qy      193  TTCGTGGACTACCCCGACCAGAACTGGGCCATCCGCGCCATCGAGACCCCTCGGGTAAA  252
             ||||||||||| |   ||||||| |||||| |||  ||||||| |||||| |||||||
Db      371  TTCGTGGACTGCCCGGACGAGAGCTGGGCCCTCAAGGCCATCGAGCGCTTTCAGTAAA  430

Qy      253  GTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAGCTAAGGAGC  312
             |||||| ||| | ||||  |||  ||||||||||||||| ||    ||| ||||| ||
Db      431  ATAGAACTGCACGGAAACCCATAGAAGTTGAGCACTCGGTCCCAAAAAAGGCAAAGGATT  490

Qy      313  AGGAAAATTCAGATTCGAAACATCCCCTCCACCTGCGAGTGGGAGTGTTGGATGGACTT  372
             ||||||||||||||| |||| |||||||||| |||||||| |||||||| |||||| ||
Db      491  CGGAAACTTCAGATACGAAATATCCCGCCCTCATTACAGTCGGGAGTGCTGGATAGTTTA  550

Qy      373  TTGGCTCAATATGGACAGTGGAGATGTGGAGAATGTGAACAGTCAACACAGACAGAACCGCC  432
             |||| || |||||| |||||||| |||| ||||||   | ||||| ||||| ||||||||||
Db      551  CTAGTCCAGTATGGAGTGGTGGAGAGCTGTGAGCAAGTGAACACTGACTCGGAAACTGCA  610

Qy      433  GTTGTCAAGTCACATATGCACAGAGAAGAAGCAAAAATAGCCATGGAAGAAGCTAAGC  492
             ||||||||||||| || ||||||| |||||||||||||| ||| || || ||| |
Db      611  GTTGTAAATGTAACCTATTCCAGTAGGACCAAGCACTAGACAAGCAAACTGAAT  670

Qy      493  GGGCATCAGTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGATGAAGAGGTGAGC  552
             ||| | ||||| |||| | |   |||||||| |||||||||||| |||||||| ||
Db      671  GGATTTCAGTTAGAGAATTTCACCTTGAAAGTAGCCTATATCCCTGATGAAATGGCCGCC  730

Qy      553  TCCCCTTCGCCCCCTCAGCGAGCCCAGCGTGGGGACCACTCTTCCCGGGACCAAGGC--  609
             |||| || |||| || |||| ||  ||| |  |||| || ||||| ||||||||  ||
Db      731  CAGCAAAACCCTTGCAGCAGCCCCGAGGTCGCCGGGGCTTGGGCAGAGGGGCTCCTCA  790

Qy      610  ------CACGGGCTCCCAGCTTCTCAGCACCAGAGATTGATTTCCCGCTGCGGATC  663
                   ||| || ||| || |||||||| | |||||| |||  ||| ||| |||||
Db      791  AGGGAGGGGTCTCCAGGATCCGTATCCAAGCAGAAACCATGGATCCGATTTGCCTCTGCGCCTG  850

Qy      664  CTGGTCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGACCATAAAGAAC  723
             ||||| |||| ||||||| |||| |   ||||||||||||  || | || ||
Db      851  CTGGTTCCACCCAATTGTTGGAGCATCATAGGAAAGAAAGAGGTGCCACCATTCGAAC  910

Qy      724  ATCACTAAGCAGACCCAGTCCCGGGTAGAATATCCATAGAAAGAGAACTCTGGAGCTGCA  783
             |||| |  | ||||| ||||||||| ||||| |   | ||||| || |  |||||| |
Db      911  ATCACCAAACAGACCCAGTCTAAAATCGATGTCACCGTAAAGAAATGCGGGGGCTGCT  970

Qy      784  GAGAAGCCTGTCACCATGCCACCCAGAGGGACTTCTGAAGCATGCCGCATGATT  843
             |||| |  | | | ||  | || |||||   || ||||  || | ||||| |
Db      971  GAGAAGTGCGATTACTATCCTCTCCTGAAGGCACCTGCCTTGAAGTCTATT  1030

Qy      844  CTTGAAATCATGCAGAAAGAGGCAGATGAGCAAGATAACAATCCTCTGAAA  903
             ||||| | ||| |||  || ||| ||||| ||| |   ||||  ||||||
Db     1031  CTGGAGATTATGCATAGGAAGCTCAAGATATAAAATTCACAGAAGAGATCCCCTTGAAG  1090

Qy      904  ATCTTGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAGATGGAAATTTGAAG  963
             ||  |   ||  |||  |||| | |   ||| ||||  |||| || || |||||
Db     1091  ATTTTAGCTCATAATAACTTTGTTGGACGCTTTATTGGTAAAGAAGGAGGAATCTTAAA  1150

Qy      964  AAAATGAACATGAAACAGGGACCAAGATAACAATCACGATTTGCAGGATTTGAGCATA  1023
             ||||| |  || || ||| ||||||||  |||  ||| || |||||| ||| ||||
Db     1151  AAAATTGAGCAAGACAGACACTAAATCACGATATCTCCATTCAGGAATTCAGGCGCTG  1210

Qy     1024  TACAACCCGGAAAGAACCATCACTGTGAAGGCCACAGTTGAGGCCTGTCCAGTGCTGAG  1083
             || | ||  ||||| || || |   | ||| | |||||| ||| |||| || ||||
Db     1211  TATAATCCAGAACGACTATTACAGTTAAGGCAATGTTGAGACATGTGCCAAAGCTGAG  1270

Qy     1084  ATAGAGATTATGAAGAAGCTGCGTGAGCCGTTGAAAATCATATCTGCCTGTGTTAACCAA  1143
             || |||  | ||| ||  | ||  |   || | |    ||  |||  |||| |||  |||
```

```
;
;  NAME/KEY: misc_feature
;  LOCATION: (1)...(4181)
;  OTHER INFORMATION: n=A,T,C or G

US-09-480-884A-175

Query Match          20.8%;   Score 710.4;  DB 4;   Length 4181;
Best Local Similarity 64.5%;   Pred. No. 2.8e-188;
Matches 1155;  Conservative   0;  Mismatches 556;  Indels 81;  Gaps 3;
```

```
Qy    73 ATGAACAAGCTTACATGGGAACCTGAGCCCGCGTCACCGGCGACGACCTCCGGCAG       132
         |||||||| ||| || ||  ||||||||| |||  | |    |   || ||||  |||
Db   251 ATGAACAAACTGTATATCCGAAACCTCAGCGAGACGCGCCCCCTCCGGACCTAGAAAGT   310

Qy   133 CTCTTTGGGGACAGGAAGCTGCCCCTGGGGGACAGGTCCTGTGAAGTCCGGCTACGCC    192
         | ||| ||||||||| |||  ||   |||| || ||| | |||||| |||||||||||
Db   311 ATCTTCAAGGACGCCAAGATCCCGTGCGGGACCCTTCCTGGTGGAGCTGCTACGCG     370

Qy   193 TTCGTGGACTACCCCGACCAGAACTGGGCCATCGGGCCTCTCCGGGTAAA           252
         ||||| |||||  ||||| || |  ||| ||  || || |  | |||||||
Db   371 TTCGTGGACTGCCCGACGAGCGCTCGGGCCCTCAAGGCCATCAGCGGCCGCCTTTCAGGTAAA   430

Qy   253 GTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGTCTAAGAGAGC  312
         ||| || |||| ||||| | ||  || || |  || ||||| |  ||  || |  || || |
Db   431 ATAGAACTGCACGGGGAAACCCATAGAGTGAGCACTCGGTCCCAAAAAAGGCAAAGGATT    490

Qy   313 AGGAAAATCAGATTCGAAACATCCCTCCACCTGCAGTGGGAGGTGTTGGATGGACTTT     372
         | ||||| |||| | |||| | ||| ||| |||| ||||||  ||| || || |  |||
Db   491 CGGAAACTTCAGATACGGAGAAATATCCCGGCTCATTTACAGTGGGGATGGATGAGTTTA   550

Qy   373 TTGGCTCAATATGGGACAGTGAGAATGTGGAACAAGTGGAACAAGAAGAGCTAAGC      432
         || || | || ||| ||| || || |||||| |||  | || || ||||||| |||
Db   551 CTAGTCCAGTGATGAGTGTGGGGAGCGTGGAGTCGTCGCAGTGAACACTGACTGGGAAACTCA   610

Qy   433 GTTGTCAACGTCACATATGCAACAAGAAGAAGAGCAAGTAGAGAGCTAGC          492
         || ||  ||  |||  |  || ||| | ||| || ||   | ||  ||| |
Db   611 GTTGTAAATGTTACCTATTCCAGTTCCGGAGAGCGTGGAGTCGTCGCAGAACAACTGAAT    670

Qy   493 GGGCATCAGTTGGAGAACTACTCCCGGAGATTTCCAAGATTCCTACAT         552
         |||  ||||| |||| |||| ||  || |||| || | | |  |  |
Db   671 GGATTTCAGTTGAGAATTCAGCCTTGAAAGTAGCCTATATCCTGATGAAATGAATGGCCGCC   730

Qy   553 TCCCCTTGCGCCCCTCAGCGAGCCCAGCGTGGGGAGCAAGGC---            609
         |||  | |||| ||||| |  | |||| |||| || |  |
Db   731 CAGCAAACCCCTTGCAGCAGCCCGAGTCGCGGGGGCTTGGGCAGAGGGGCTCCTCA        790

Qy   610 ------CACGGCCCTGGGGCCACTTCTCAGGCCAGAAATACCCTCAGGGAGTC        663
               |||||||||| | |  |  | || |||||||  | ||| | |  ||||
Db   791 AGGCAGGGGTCTCCAGGATCCGTATCCAGCCAGCAGAAACCATGTGATTTGCCTCTGCGCCCTG   850

Qy   664 CTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGAAC                723
         ||||||||||||| ||| ||| ||| ||| ||| ||||||
Db   851 CTGGTTCCCCACCCAATTTGTTGGAGCCATCATCATGATAAAGAAGTGCACCATTCGGAAC   910

Qy   724 ATCACTAAGCACAGACCCCAGTCCCGGTAGTATCCCATAGGCACTCTGGAGCTGCA        783
         || ||| |||| | || |||| | ||  | || ||  |  |  |  || |||||||
Db   911 ATCACCAAACAGACCCAGTCCTCAAAATCGATGTCCACGCTAAAGAAAATGCGGGGGCTGCT    970

Qy   784 GAGAAGCCGTGTCACCATCCATGCCACCCCAGAGGGACTTCTGAAGCATGCCGCATGATT     843
         |||||||| || ||||| || ||||| |||||| ||   | |||||||| || ||||||
Db   971 GAGAGTGCGATTACTACTTGGTGACGTCTTATTGGTGGAGTCGTGTAAGTCTATT        1030

Qy   844 CTTGAAATCATGCAGAGAAGAGCAGCAGTGAGACCAACTAGCCCGAAGATTCCTCTGAAA      903
         |||| ||||  | |  |||||| |||  | ||||| | || |  |  | | ||| || |
Db  1031 CTGGAGATTATGCATACAAGGAAGAAGTCAAGATAATAAAATCACAGAAAGAGATCCCCTTGAAG  1090

Qy   904 ATCTTGGCACACAAATGGCTTGGTTGGAAGAGCTAAAATCAACTTTGAAGG            963
         || |||  | ||| |  || ||   ||||  || | | | | ||| || |
Db  1091 ATTTTAGCTCATAAATACTTGTTGGACGTGTCTTATTTGTGTAAAGAAGAAGAATCTTAAA   1150

Qy   964 AAAAATTGAACATGAAACAGGGACCAAGATAACAATCCATCTTTGAGCATA             1023
         |||| || |||| || ||| ||| |||| | | ||| || | || |||
Db  1151 AAAATTGAGCAAGACACAGACACTAAAAATCCGATATCCATTGCAGACGACGCTG         1210
```

```
Db  1271 GAGGAGATCATGAAGAAATCAGGGAGTCTTATGAAAAATGATATTGCTTCTATGAATCTT   1330
         |||||| |||| ||| |  | |   | | |   ||| ||| ||| |||| |  | ||||
Qy  1144 CAAGCCAATCGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTCAACAGGACTG     1203

Db  1331 CAAGCACATTTAATTCCTGGATTAAATCTGAACGGCTTGGGTCTGTTCCCACCCACTTCA   1390
         |||||||||| |||||||||||| |||||||| ||||||| |||| |||||||  |||
Qy  1204 CAAGCACATTTAATTCCTGATTAAATCTGAACGGCTTGGGTCTGTTCCCACCCACCACCG    1263

Db  1391 GGGATGCCACCTCCCACCTCAGGGCCCCCTTCAGCCATGACTCCT-----------     1435
         |||||||||||||||  | |||||||||| ||||||||  ||||
Qy  1264 TTCACTACCCCTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTGGCCGGTTC     1323

Db  1436 --------------------CCCTACCCGCAGTTTGAGC------              1454
                             |  || || ||| ||
Qy  1324 CCGCATCATCACTCTTATCCAGACCAGGAGAATCTCTTCATCCCCAACCCAGGCT       1383

Db  1455 -------AATCAGAAACGGACACTGTTCATCAGTTTATCCCAGCTCATCA           1498
                |||||| || || ||| || ||||||| |||  | ||| |  | |
Qy  1384 GTGGGGCCGCCATCATCGGGAAGAAGGGGGCACACATCAAACAGCTGGCGAGATTCGCCGGA  1443

Db  1499 GTCGGTGCGCCATCATCGGCAAGCAGGGCCAGCACATCTTCTCGCTTTGCTGGA        1558
         || ||||| |||||||||| || |  || || |||||| | |||  || ||||
Qy  1444 GCCTCTATCAAGATGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCATC     1503

Db  1559 GCTTCAATTAAGATTGCTCCAGCCGAAGCACCAGATGCTAAAGTGAGGATGGTGATTATC   1618
         || || || |||||| | ||||||||| || ||||| ||| || ||||||||| |||||
Qy  1504 ACCGGGCCACCGGAGCCCAGTTCAAGGCCCAGGACGGATCTTGGGAAACTGAAAGAG      1563

Db  1619 ACTGGACCACCAGAGCTCAGTTCAAGGCTTCAAGGGCAAGAATTTATGGAAAAATAAAGAA  1678
         || || ||||| ||| |||||||||| | || || | ||| |||| |||||| | |||||
Qy  1564 GAAAACTTCTTAACCCCAAAGAGAAGAAGTGAAGCGATATCAGAGTGCCCTCT         1623

Db  1679 ACTGGACCACCAGAGCTCAGTTCAAGGCTGAAAGTTATGGAAAAAATAAAGAA          1678
         || || |||| |||| ||||||||| | ||| |  | ||||||||| ||||
Qy  1564 GAAAACTTCTTAACCCCAAAGAGAAGAAGTGAAGCGATATCAGAGTGCCCTCT         1623

Db  1739 TTTGCTGTGCAGAGTTAGTGTTCCTAAAGAACTTGAGCTGAATGAACTTCAGAATTTG    1798
         || || ||||||||| | ||| |  |||||| | | || |  |||| |  |||||||||
Qy  1624 TCCACAGCTCGGCCGGGTGATTGGGCAGGTGGCAAGTGGCAAGACCCTGAACGAACTTA    1683

Db  1799 TCAAGTGCAGAAGTTGTGTCCCTCGTGACACACTTGATGAGAATGACCAAGTGGTT       1858
         | |||||||||| | | || |  | |||||||| || | |  |  |   | || ||
Qy  1684 ACCAGTGCAGAAGTCATCGTCGTCCGTGACCAGAGATGAAAATGAGGAGAAGTGATC       1743

Db  1859 GTCAAAATAAACTGGTCACTTCTATGCTTCCAGGTTGCCCCAGAGAAAAATTCAGGAAATT   1918
         || ||||| ||||| || ||  | |  |||| | | |||||| |||| ||||| |||||
Qy  1744 GTCAGAATTATCGGGCACTTCTTTGTGCTAGCGACGTCGACAGATCAGGGAAATT        1803

Db  1919 CTGACTCAGTGAAAGCAGCACCAACAACAGAGGGCTCTGCAAGTGGACCAC            1970
         ||||| |||| |||||||| | ||  | |  || ||| | ||| ||  |||
Qy  1804 GTACAACAGTGAAGCTGGAAGCAGCAGCACCAACAACAGAGGAGCTCCAGGGGTCTGCCCTCAC  1855
```

```
RESULT 11
US-09-480-884A-175
;  Sequence 175, Application US/09480884A
;  Patent No. 6482597
;  GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
;  APPLICANT: Fan, Liqun
;  APPLICANT: Hosken, Nancy A.
;  APPLICANT: Kalos, Michael D.
;  APPLICANT: Fanger, Gary R.
;  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
;  TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 2101121.455C6
;  CURRENT APPLICATION NUMBER: US/09/480,884A
;  CURRENT FILING DATE: 2001-08-27
;  NUMBER OF SEQ ID NOS: 330
;  SOFTWARE: FastSEQ for Windows Version 3.0
;  SEQ ID NO 175
;  LENGTH: 4181
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
```

```
NUMBER OF SEQ ID NOS:  350
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 175
LENGTH: 4181
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: (3347)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (3502)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (3506)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (3520)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (3538)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (3549)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (3646)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (3940)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (3968)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (3974)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (4036)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (4056)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (4062)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (4080)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (4088)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (4115)
OTHER INFORMATION: n=A,T,C or G
US-09-542-615A-175
```

```
Query Match             20.8%;  Score 710.4; DB 4;  Length 4181;
Best Local Similarity   64.5%;  Pred. No. 2.8e-188;
Matches 1155;  Conservative   0;  Mismatches 556;  Indels   81;  Gaps    3;

Qy    73 ATGAACAAGCTTTACATCGGGAACTTCAGCCCGTCACCGCCGCAGCCTCCGGCAG 132
         ||||||     |||||||||||  || |||||  ||||| ||| |||||  |||||
Db   251 ATGAACAAACTGTATATCGGAAACCTCAGCGAGACGCCGCCCCTCGGGACCTAGAAAGT 310

Qy   133 CTCTTTGGGGACAGGAAGCTGCCCCTGGCGGGACAGGTCCTGCTGAAGTCCGGCTACGCC 192
         |||||||| ||||| |||||||||    ||| ||||| ||||  ||||||  |||||||||
Db   311 ATCTTCAAGAGACGCCAAGATCCGGTGTCGGGACCCTTCCTGGTGAAGACTGGCTACGCG 370

Qy   193 TTCGTGGACTACCCCGACCAGAACTGGGCCATCGCCGCCAGAGCCCTCGGGTAAA 252
         |||||||| ||||| |  |||| ||||| |||||||  ||| |  ||  |||||||
Db   371 TTCGTGGACTGCCCCGGACGAGAGCTGGGCCCTCAAGGCCATCGAGGCGCTTTCAGTGTAAA 430
```

```
RESULT 12
US-09-542-615A-175
; Sequence 175, Application US/09542615A
; Patent No. 6518256
GENERAL INFORMATION:
APPLICANT:  Wang, Tongtong
APPLICANT:  Fan, Liqun
APPLICANT:  Kalos, Michael D.
APPLICANT:  Bangur, Chaitanya S.
APPLICANT:  Hosken, Nancy A.
APPLICANT:  Fanger, Gary R.
TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THERAPY
TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE:  210121.455C8
CURRENT APPLICATION NUMBER: US/09/542,615A
CURRENT FILING DATE: 2000-04-14
```

```
Qy  1024 TACAACCCGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGAG 1083
          |||||  || ||||||  || ||||||||||  ||||  |||| || || ||||||||||
Db  1211 TATAATCCAGAACGCACTATTACAGTTAAAGGCAATGTTGAGACATGTGCCAAAGCTGAG 1270

Qy  1084 ATGAGATTATGAGAAGCTGCGTGAGCGCCTTTGAAAATGATATGCTGGCTGTTAACCAA 1143
          |||| |||| ||||||   || ||||||||||||||||||||| ||||||| ||||||
Db  1271 GAGGAGATCATGAGAAAATCAGGAGTCTTATGAAAATGATATTGCTTCTATGAATCTT 1330

Qy  1144 CAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGACTG 1203
          |||| |||| | ||| ||||| || |  ||||||| ||||| | |||||||||| || |
Db  1331 CAAGCACATTTAATTCCTGGATTAAATTGAACGGCCTTGGGTCTGTTCCCACCCACTTCA 1390

Qy  1204 TCCGTGCTATCTCCACCAGCAGCGGGCCCCCGGAGCTCCCCCGCTGCCCCTACCACCCC 1263
          |||| |||| | |||| ||||| ||| | || || ||| || |  ||| |||
Db  1391 GGGATGCCACCTCCCACCTCCAGGGCCCCCTTCAGCCATGACTCCT------ 1435

Qy  1264 TTCACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTGCCCGTTC 1323
                                                  ----CCCTACCCGCAGTTGAGC------
Db  1436 ------ 1454

Qy  1324 CCGCATCATCACTCTTATCCAGAGCAGAGATTGTGAATCTCTTCATCCACCCAGCT 1383
          ------          ||||  ||||  ||| ||  | ||||| ||| |||||||
Db  1455 ------AATCAGAAACGGAGACTGTTCATCAGTTTATCCCAGCTCTATCA 1498

Qy  1384 GTGGGCGCCATCATCGGGAAGAAGGGGCCACACATCAAACAGCTGGCCAGATTCGCCGGA 1443
          ||||| ||||||||| ||||| |||  || |||||||| ||||  | |||||| |||||
Db  1499 GTCGTGCCATCATCGGCACAGGCCAGCAGGCCGCACTTTTCGCTTTGCTGGA 1558

Qy  1444 GCCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCATC 1503
          ||  ||||||||| |||| |    ||||| ||||| || | ||| |||| ||| ||||
Db  1559 GCTTCAATTAAGATTGCTCCAGCGGAAGCACCAGATGCTAAAGTGAGGATGGTGATTATC 1618

Qy  1504 ACCGGGCCACCGGAAGCCCAGTTCAAGGCCAGGACCGATCTTTGGGAAACTGAAAGAG 1563
          |||| | |||| | ||| ||||  |||| || ||  |   || |  |||   ||| |||
Db  1619 ACTGACCACCAGGAGGCTCAGTTCAAGGCTCAGGAAGAATTTATGAAAAATTAAAGAA 1678

Qy  1564 GAAAACTTCTTTAACCCCAAAGAAGAGAAGGTGAAGCTGAAGCGCCATATCAGAGTGCCCTCT 1623
          ||||  ||||| | |  |||||| | | || |  |  |||||| || | | |||||| ||||
Db  1679 GAAAACTTTGTTAGTCCTAAAGAAGAGGGTGAAACTTGAAGCTCATATCATCAGTAGTGCCATCC 1738

Qy  1624 TCCACAGCTGGCCGGGTGATTGGCAAAAGTGGCAAGACCGTGAACGAACTGCAGAACTTA 1683
          | || | ||| |  | ||||| ||||| | || ||| |  || ||| || |||| || |
Db  1739 TTTGCTGCTGGCAGAGTTATTGGAAAAAGGAGGCAAAACGGTGAATGACCAAGTGTTT 1798

Qy  1684 ACCAGTGCAGAGTCATCGTGCCTCGTTGCCAACAAACGCCAGATGAAAATGAGGAAGTGATC 1743
          || || | || |||| | || |||| ||| ||  |||| | || |||| ||||| |||||
Db  1799 TCAAGTGCAGAAGTTGTTGTCCCGTGAACACACCTGATCGATGATGCCAAGTGTT 1858

Qy  1744 GTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAATT 1803
          ||| ||||| | ||| |  ||  ||| | |  || || ||| |  ||  |||| ||| |
Db  1859 GTCAAAATAACTGGTCACTTCTATGCTTGCCCAGAGAAAAATTCAGGAAATT 1918

Qy  1804 GTACACAGTGAAGCAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCAC 1855
          |||||        |||| ||| |||| || | || || |  |||  |||||||
Db  1919 CTGACTCAGTAAAGCAGCACCAACAGACAGAGAAGCGCTTCAAGGCTCTGGGTGGGACCAC 1970
```

Qy 253 GTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGGAGC 312
         |||||| ||||||||| |||||| ||| || |||||||| | |||||||||||||||
Db 431 ATAGAACTGCACGGGAAAACCATAGAAGTTGAGCACTCCGGTCCCAAAAAGGCAAAGGATT 490

Qy 313 AGGAAAAATTCAGATTCGAAACATCCTCCTCACCTGCAGTGGAGGTGTTGGATGACTT 372
         ||||| |||| ||||| |||||| ||||| ||||||| ||||||||||| |||||||
Db 491 CGGAAAACTTCAGATACGAGAAATATCCCGCCTCATTTACAGTGGGAGGTGCTGGATAGTTTA 550

Qy 373 TTGGCTCAATATGGGACAGTGGAGAATGTGAACAAGTCAACACAGACACAGAAACCGCC 432
         ||||||||||||| ||||||||||| ||| ||||||| || | ||||||||| |||||
Db 551 CTAGTCCAGTATGGAGTGGTGGAGAGCTGTGAGCAAGTGAACACTGCTAAAGTGGTGATTATC 610

Qy 433 GTTGTCAAGTCACATATGCAACAAGAGAAGAAGCAAAAATAGCCATGGAGAAGCTAAGC 492
         |||||||||| |||| ||| ||||| ||||| ||||||| ||||| |||||||| |||
Db 611 GTTGTAAATGTAACCTATTCCAGTAAGGACCAAGTTAGACAACAAACTGAAT 670

Qy 493 GGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCTACATCCCGGATGAAGAGGTGAGC 552
         ||||| |||| |||||||| ||||| |||||| |||||| |||| | |||||||||
Db 671 GGATTTCAGTTAGAGAATTTCACCTTGAAAGTAGCCTATAATCCCTGATGAAATGGCCGCC 730

Qy 553 TCCCCTTCGCCCCCTCAGCGAGCCCAGCGTGGGACCACTCTTCCCGGGACCAAGGC--- 609
         |||| | | ||||| || || |||| ||||| || | ||||||||
Db 731 CAGCAAAACCCCTGCAGCGAGCCCGGAGGTCGCCCGGGGGTTGGGCAGAGGGGCTCCTCA 790

Qy 610 -----CACGGCCCTGGGGCACTTCTCAGCCAGATGATTCCCGTCGCGGATC 663
              | | |||| | ||| ||||| ||||| |||| || ||| |||||
Db 791 AGGCAGGGGTCTCCAGGATCCGTATCCCAAGCAGAGAAACCATGTGATTTGAATTTGCCTCTGCCGCCTG 850

Qy 664 CTGGTCCCCACCCAGTTTGTTGGTGTCCATCATCGGAAAGGAGGGCTTGACCATAAAGAAC 723
         |||||| ||||||| ||||| || || || |||||| || | |||||||| |||||
Db 851 CTGGTTCCCAACCAATTGTTGGAGGCATCATAGGAAAAGAGAAGGTGCCACCATTGGAAC 910

Qy 724 ATCACTAAGCAGACCCCAGTCCCGGGTGTAGTATCCATAGAAAAGAGAACTCTGGAGCTGCA 783
         |||| |||||| |||| |||| || | ||| || |||||||||||| |||||||
Db 911 ATCACCAAACAGACACCCAGTCCCGGGTAAAATCGATGTCACCGTAAAGAAAATGCGGGGCTGCT 970

Qy 784 GAGAAGCCTGTCACCATCCATGCCACCCAGAGGGACTTCTGAAGCATGCCGCATGATT 843
         |||||| |||| |||| ||| ||||| ||||||||| ||||||| ||||| |||||
Db 971 GAGAAGTGATTACTATCCTTCTACTCCTGAAGGGCACCTCTGAAGGCTTGTAAGTCTATT 1030

Qy 844 CTTGAAATCATGCAGAAGAGAGGCAGATGAGACCGAAGAGATTCCTCTGAAA 903
         |||||| || || |||||||||| || ||||||| ||| | ||| || |||
Db 1031 CTGGAGATTATGCATAAGGAAGCTCAAGATCTTGTTGACGTCTTATTGGTAAAGAGGAAGAAGATCCCCCTGAAG 1090

Qy 904 ATCTTGGCACACAATGGCTTGGTTGGAGACTGATTGGAAAGAGGCAGAAATTGAAG 963
         || | | || | ||| | ||| ||||| || || ||||| || |||
Db 1091 ATTTAGCTCATAATAACTTTGTTGACGTCTTATTGGTAAAGAAGGAGAAGAAATCTTAAA 1150

Qy 964 AAAATTGAACATGAAACATGGAGACCAAGATAACAATCCATCTTTGCAGGATTTGAGCATA 1023
         |||||| ||| || ||| || ||||||| ||||| ||| || ||||||| |||| |||
Db 1151 AAAATTGAGCAGAGACACAGACAGAATCTAAAATCACGATATCTCCATTGCAGGAATTGCAGCGCTG 1210

Qy 1024 TACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGAG 1083
         ||| | || ||| | || |||||| |||| | ||| | |||||| |||| || || |
Db 1211 TATAATCCCAGAACCGCACTATTACAGTTAAAGGCAATGTTGAGACATGTTGAGAAGCTGAG 1270

Qy 1084 ATAGAGATTATGAAGAAGCTGCGTGAGGGCCTTTGAAAATGATATGCTGGCTGTTAACCAA 1143
         || | | | || |||| || || ||||| |||| |||| | | ||| ||| ||||||||
Db 1271 GAGGAGATCATGAAGAAATCAGGAGTCTTATGTGAAAAAATGATATTGCTTCTATGGAATCTT 1330

Qy 1144 CAAGCCAATCTGATCCCAGGGTTGAACCTTGGCATCTTTTCAACAGGACTG 1203
         |||||| |||| || || | ||| |||| | || ||||| | ||| |||
Db 1331 CAAGCACATTTAATTCCTGGATTAAAATCTGAACGCCTTGGGTCTGTTGTCCCCACCCACTTCA 1390

Qy 1204 TCCGTGCTATCTCCACCAGCAGGGCCCCCGGAGCTCCCCGCTGCCCCCTACCACCCC 1263
         |||| ||| | ||||||||| |||| | ||| | || ||| | |||
Db 1391 GGGATGCCACCTCCCACCTCCAGGGTCCCCCCTTCAGCCATGACTCCT----------- 1435

Qy 1264 TTCACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTTGGCCCGTTC 1323
         | | ||| | ||||| ||| ||||||
Db 1436 -------------------------CCCTACCCCGCAGTTTGAGC------ 1454

Qy 1324 CCGACATCATCACTCTTATCCAGAGCAGGAGGAGATTGTGAATCTCTTCATCCCAACCCAGGCT 1383

Db 1455 --------------AATCAGAAACGGAGGACAGCTGTTCATCAGTTATCCAGCAGTTATCA 1498

Qy 1384 GTGGGGGCCATCATCGGGAAGAAGGGGGCACACATCAAACAGCTGGCCGATTGGCCGGA 1443
         |||| | |||||| ||| ||||||| ||| || |||||||| || ||| || ||| ||
Db 1499 GTCGGGTGCCATCATCGGCAACAGGGGCGAGGCCGGAGGTGCCGGAGGCCTTTCTCGCTTTGCTGGA 1558

Qy 1444 GCCTCTATCAAGATGCCCGTGCCCCTGCGAAGGCCCAGCGTCGTGCGAAAGGATGGTCATCATC 1503
         ||| | | || || | ||||| ||| |||| |||| | || || || || | ||| ||
Db 1559 GCTTCAATTAAGATTGCTCAAGCGGAAGCACCAGATGCTAAAGTGCTAAAGTGAGGTGATTATC 1618

Qy 1504 ACCGGGCCACCGGAAGCCCAGTTCAAGGGCCCAGGGACGATCTTTGGAAACTGAAAGAG 1563
         || ||| | |||||||| || ||||| || |||| |||| | | ||| ||| | |||
Db 1619 ACTGGACCACCAGAGGCTGTTCAGGTTCAAGGCTCAGGGAAGAATTTATGGAAAAATTAAAGAA 1678

Qy 1564 GAAAACTTCTTAACCCCAAAGAGAGAAGTGAAGCTGAAGCGCATATCAGAGTGCCCTCT 1623
         |||||| | ||| || |||| |||| |||||| ||||||| |||| |||||||| ||
Db 1679 GAAAACTTTGTTAGTCCTAAAGAAGAGGTGAAACTTGAAGTCATATCAGAGTGCCATCC 1738

Qy 1624 TCCAGAGCTGGCCGGGTGATTGGCAAGATGGTGCAAGAGTGCAAGACGCACGAACTTA 1683
         ||| |||| |||| || || || || | |||||| ||||| || |||||| |||
Db 1739 TTTGCTGCTGGCAGAGTTATTGGAAAAGGAGGCAAAAACGTGAATGAACTTCAGAATTTG 1798

Qy 1684 ACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAAGCGCCAGATGAAAATGAAGGAGTGATC 1743
         ||| || ||||| | || || || ||| |||| | || |||| |||| |||||||| ||
Db 1799 TCAAGTGCAGAAGTGTGTCCCTCGTGACCAGCACCTGATGACCAACGTGGTT 1858

Qy 1744 GTCAGAAT TATCGGGCACTCTTCTTGGTGTCTAGCCAGGACTTGCACAGCGCAAGATCAGGGAAATT 1803
         || |||| | || || ||| |||||| |||| || |||| | | || ||||||| |||||
Db 1859 GTCAAAATAACTGGTCACTTCTATGCTTGCCAGTTGCCCAGAGAAAATTCAGGAAATT 1918

Qy 1804 GTACAACAGGTGAAGCAGCAGCAGGAGCAGAAATACCCTCAGGGAGTGCGCCTCAC 1855

Db 1919 CTGACTCAGTAAAGCAGCAGCACCACAACAGAGAAGGCTCTGCAAAGTGGACCAC 1970

```
FEATURE:
NAME/KEY: unsure
LOCATION: (3520)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (3538)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (3549)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (3646)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (3940)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (3968)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (3974)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (4036)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (4056)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (4062)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (4080)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (4088)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (4115)
OTHER INFORMATION: n=A,T,C or G
US-09-606-421B-175

Query Match        20.8%;
Best Local Similarity  64.5%;
Matches 1155; Conservative   0;
```

```
Qy    313 AGGAAAATTCAGATTCGAGAACATCCCTCTCACCTGCAGTGGGAGGTGTTGGATGGACTT 372
                ||||||||  |    |   |||||||   |||||||||  ||||| ||||
Db    491 CGGAAACTTCAGATACGAGAAATATCCCGCCTCATTTACAGTGGGAGGTGCTGGATAGTTTA 550

Qy    373 TTGGCTCAAATATGGACAGTGGAGAATGTGGAGAGCTGTGAGCAAGTCAACAGAACAAGCGCC 432
               |||      |||||| ||||||||| ||||| |||| |||  |||||||  || ||
Db    551 CTAGTCCAGTATGGAGTGGTGTGAGCAAGTGAGCAAGTGAACACTGACTCCGAAACTGCA 610

Qy    433 GTTGTCAAGCTCACATATGCAACAAGAGAGAAAAATAGCCATGGAGAAGCTAAGC 492
               |||||||    | | |||| |||  ||||| | |||   || |||  |||||
Db    611 GTTGTAAATGTAACCTATTCCAGTAAGGACCAAGCTAGACAAGCACTAGACAAAACTGAAT 670

Qy    493 GGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGATGAAGAGGTGAGC 552
               |||||||||||| ||||   |||   ||| ||||| |   ||  ||| ||||| ||||
Db    671 GGATTTCAGTTAGAGAATTTCACCTTGAAAGTAGCCTATATCCCTGATGAAATGGCCGCC 730

Qy    553 TCCCCTTCGCCCCTCAGCGAGCCCCAGCGTCTGGGGACCACTCTTCCCGGGAGCAAGGC--- 609
               |||||||||   ||     ||||   |   ||||   ||||||||||||||| ||||
Db    731 CAGCAAAACCCTTGCAGCAGCCCGAGGTCGCCGGGGCTTGGGCAGAGGGGCTCCTCA 790

Qy    610 ------CACGGCCCCTGGGGCACTTCTCAGCGCAGACAGATTGATTTCCCGTGCCGGATC 663
                     ||||||   ||  ||||| ||||||| ||||| | ||||| || ||||||
Db    791 AGGCAGGGGGTCTTCCAGATCCGTATCCAAGCAGAAACCAATGTGATTTGCCTCTGGGCCCTG 850

Qy    664 CTGGTCCCCACCAGTTTGTTGGTCCATCATGCGGATGGGCGTTGACCATAAAGAAC 723
               |||||||| ||||||||||||  ||| |||    |||||| |||  ||||| |||
Db    851 CTGGTTCCCACCCAATTTGTTGAGCCATCATAGGGAAAAGAGGTGCCACCATTCCGAAC 910

Qy    724 ATCACTAAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAGAGAACTCTGGAGCTGCA 783
               |||||||||||||||||||||||  ||| |||  |||||| || ||||| ||||||
Db    911 ATCACCAAACAGACCCAGTCTCAAAATCGATGTCCACGTAAAGAAAATGGGGGGCTGCT 970

Qy    784 GAGAAGCTGTCACCATCCATGCCAGAGGGACTTCTGAAGCATGCCGCATGATT 843
               |||||| ||||| | | ||| ||||||| |||||||| || || ||| ||||
Db    971 GAGAAGTCGATTACTATCCTGAAGGCCACCTCTGCGGCTTGTAAGTCTATT 1030

Qy    844 CTTGAAATCATGCAGAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAA 903
               |||||| | ||||||| |||| ||| ||| | | || ||||| ||||||| ||||||
Db    1031 CTGGAGATTATGCATAAGGAAGAGCTCAAGATATAAAATTCACAGAAGATCCCCTTGAAG 1090

Qy    904 ATCTTGGGCACACAATGGCTTGTGTTGAAGACTGATTGGAAAGAAGCCAGAAATTTGAAG 963
               ||| |   | | |||  |  ||||||| || ||| |||| |||| ||||| ||||||
Db    1091 ATTTTAGTCATGATAAACTTTGTTGGACGTCTTATTGGTAAAGAAGAAATCTTAAA 1150

Qy    964 AAAATTGAACATGAAACAGGGACCAAGATAACAATCCATCTTTGCAGGATTTGAGCCATA 1023
               |||||||||||||   ||||||||| | |  | |||||||| ||||||  || |||||
Db    1151 AAAATTGACAAGACACAGACTAAAATCACGATATCTCCAGCTGGCGAGGAATTGACGCTG 1210

Qy    1024 TACAACCCGGAAAGAACCATCACTGTGGAAGGCCATGGCCTGTGCCAGTGCTGAG 1083
                |   | |||||| || ||||||  ||||| ||| |||| | ||| |||| |||
Db    1211 TATAATCCAGAACGCCACTATTACAGTTAAAGGCAATGTTGAGACAATGTGCCAAAGTCGAG 1270

Qy    1084 ATAGAGATTATGAGAAGCTGCCTGAGCCTTTGAAAATGATATGCTGGCTGTTAACCAA 1143
                ||||||||||||||||||    ||| ||| ||||||| ||| || |||| || ||||
Db    1271 GAGGAGGATCATGAAGAAAATCAGGGGAGTCTTATGAAAATGAATATTGCTTCTATGAATCTT 1330

Qy    1144 CAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGACTG 1203
               |||||| | |  ||  |||||||||||  |||  ||||||  || | ||||||| ||||
Db    1331 CAAGCACATTTAATTCCTGGATTTAAATCGAACGGCTTGGGTCGTTCCCACCCACTTCA 1390

Qy    1204 TCCGTGCTATCTCCACCAGCAGGGCCCCGGAGTCCGGCTACGCC 1263
               ||| |||| | ||||| |||| |||||||| | ||   ||||| |
Db    1391 GGGATGCCACCTCCACGGCGCCCTTCAGCCATGAGCTCCT------------ 1435

Qy    1264 TTCACTACCCACTCTTATCCGACAGAGATTCTCCAGCCTGTACCCCCATCACCAGTTTGGCCGGCTAAA 1323
                ||| ||| |||  || | ||||||||||| |||||| ||||  |||||| |||||| | ||||||
Db    1436 -------------------------------CCCTACCCGCAGTTTGAGC----- 1454

Qy    1324 CCGCATCATCACTCTTATCCGACAGAGATTGTGAATCTTCATCCCAACCCAGGCT 1383
                                                 |||| || |||| || ||||||
Db    1455 ------------------AATCAGAAACGGAGACTGTTCATCAGTTTATCCCGTCTATCA 1498
```

```
Qy     73 ATGAACAAGCTTTACATGGGACCTTGAGCCCGTCACCGCCGACCTTCCGGCAG 132
             |||||||||| || | ||||||| ||||||||||  ||||   |  |||||
Db    251 ATGAACAAACTGTATATCGGAAACCTCAGCGAGACGCGGCCCCTCAGCGACCTTAGAAGT 310

Qy    133 CTCTTTGGGACAGGAAGCTGCCCCGGGCGGGACAGGTCGCCCGGACGGTCGTGAAGTCGGCTACGCC 192
             |||   | |||| | |   ||||||| ||||||||||||  |||||||  |||||| ||||||||
Db    311 ATCTTCAAGGAGCGCGCCAAGAATCCGGGTGTCGGGACCCCTTCCTGGTGAAGACCCTCTCGGGTAAA 370

Qy    193 TTCGTGGACTACCCCGGACCAGGAACTGGGCAATCATGGGGAAAATCATGGAAGTGATTACTCAGT 252
             |||||||| ||| | |||| |||  |||| ||||||||| ||||||||  |||||||
Db    371 TTCGTGGACTGCCCGGACGGAGCTGGGCCCTCAGCGAGAGATTGTGAATCTTCATCCCAACCCAGGCT 430

Qy    253 GTGGAATTGCACGGGAAAATCATGGAGAAGTTGAGCACTCGGTCCCAAAAAGGCAAAGGATT 312
             ||||||| ||||||||  |||| ||||  |||||||||||| ||   |||||| |||||||
Db    431 ATAGAACTGCACGGGAAACCCATAGAAGTTGAGCACTCGTTCATCAGTTTATCCGCTCATCA 490

Length 4181;
Score 710.4;  DB 4;
Pred. No. 2.8e-188;
Mismatches 556; Indels  81; Gaps   3;
```

```
Qy   1384 GTGGGGCCCATCATCGGAAGAAGGGGCACACATCAAACAGCTGGCGAGATTGCCCGGA 1443
          ||||| ||| ||||||||||||| |||  |||  |||||   ||||| ||||||| ||||
Db    371 CGGAAATACAGATCCGCAATATTCCACCTCAGCTCCGATGGGAGTGCTAGATAGCCTG 430

Qy   1499 GTCGGTGCCATCATCGGCGCAGCAGGCGCCAGCAATCAAGCAGCTTTCTCGCTTGCTGGA 1558
          || ||| |||||||| ||| || ||| ||||| || || |||  ||| |  ||  | |||
Db    373 TTGGCTCAATATGGCACAGTGGAGAATGTGAGAAGGATGGTCATCATC 432

Qy   1444 GCCTCTATCAAGATTGCTCCTGGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCATC 1503
          |||| |  || ||||| | ||| |||| ||| | | ||||  |||||| ||||| |||
Db    431 CTGGCTCAGTACGTGACAGTGGAGAACTGTGAGCAGTGAACACTGAAAGTGAGACAGCG 490

Qy   1559 GCTTCAATTAAGATTGCTCACTTGCTTCATTTCTCAGGGAAGGTGGTGATTATC 1618
          ||||| ||| |||||||| ||| ||  ||  | |||| || ||  | ||  |||
Db    433 GTTGTCAACATGCAACATGACAGAGAGAGAAGCAAAATAGCCATGGAGAAGTAAGC 492

Qy   1504 ACCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGACGGATCTCTTGGGAAACTGAAAGAG 1563
          || ||| |||| ||||| ||| | ||| ||  |||| | ||| ||||| |||| |
Db    491 GTGGTCAACGTCACCTACTCTAACCGGGAGCAGAGACCAGGCAAGTCATCATGAAGCTAAAT 550

Qy   1619 ACTGGACCACCAGAGGCTCAGTTCAAGGCTCAGGGAGAATTTATGCGAAAAATTAAAGAA 1678
          |  ||| |||| ||||||  || | |||| || ||||| ||  |  |||| |  ||||
Db    493 GGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTCAGATTCCTACATCCCGGATGAAGAGGTGAGC 552

Qy   1564 GAAAACTTCTTTAAACCCAAGAGGAGTGAAGCTGGAAGCGCATATCAGAGTGCCCTCT 1623
          |||| |||||||  || | |||||  || |||| ||||| ||| ||||||||  ||
Db    551 GGCCATCAACTGGAGAACCATGCCCTGAAGGTCTCCTACATACCTGATGAGCAGATAACA 610

Qy   1679 GAAAACTTTGTTAGTCCTAAAGAAGGTGAAACTTGAAGCTCATATCAGAGTGCCATCC 1738
          |||||||   ||||||  |||  || |  |||| ||| || || ||||| ||  ||||
Db    553 TCCCCTTC--------GCCCCTCAGCGAGCGGTGGGACCCACTCTTCCCGGAG 603

Qy   1624 TCCACAGCTGGCCGGTGATTGGCAAGGTGCAAGACCGTGAACGAACTGCAGAACTTA 1683
          ||| ||||| ||| | ||| |||| | || ||| | || |||| |  ||||| ||||
Db    611 CAAGGTCCTGAGAATGGGCGTCGTGGAGCTTTGGGTCTCGGGGCCCAGCCCGGCAAGGG 670

Qy   1739 TTTGCTGCTGGCAGAGTTATTGGAAAAGGAGGCAAAAACGGCACAGTTGCTGTGAGTC 1798
          ||||| || ||| |||||  |  |||  |||| ||||| ||||| || | ||||  ||
Db    604 CAAGGCCACGCCCCTGGGGGCACTTCTCAGCCACAGATTGATTCCCGCTCGCGGATC 663

Qy   1684 ACCAGTGCGAAGTCATCGTGCCTCGTGCTGCCAAAACGCCAGATGAAAATGAAGTGATC 1743
          ||||||| ||| | || ||| ||  || ||||  || | |||| | ||||| |||| |
Db    671 TCGCCCGTGCGAGCAGGGGCTCCAGCCAGCAGCCTATCATTGGCGATGTGCTCCGCGTC 730

Qy   1799 TCAAGTGCAGAAGTGTGTTGTCCCCTGCCAGACACCTGCAGGACACCAAGCCAAGTGGTT 1858
          ||||||||||| | || || ||| ||  || | ||| || | ||||| |  | |||| ||
Db    731 CTGGTGCCTACGCAGTATGGTAGGCGGCTATCATTGGCAAGGAGGTGCCACCATCCGAAAC 790

Qy   1744 GTCAGAATTATCGGGCACTTCTTTGCTAGCCAGATGCACAGCGCAAGATCAGGGAAATT 1803
          | |||| ||||| ||| || || ||||  ||||| ||||| |  ||||| | |||| ||
Db    791 ATCACAAAACAGACCCAGTCCAAAATAGACGTGCATAGGAAGAAATGCATAGGAGGAGAATCGGGGCCTGCG 850

Qy   1859 GTCAAAATAACTGGTCACTTCTACTTCTATGCTTGCCCAGGTTGCCCAGAGAGAAAATTCAGGAAATT 1918
          || || ||| ||  ||||| |  | ||| |  |||  || ||||||| | || |  ||| ||||| ||
Db    724 ATCACTAAGCAGACCCAGTCCGGTAGTATATCCAGAGAGAAAAGAGAACTCTGAGAGCTGCA 783

Qy   1804 GTACACAGGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTGCCTCAC 1855
          ||||| ||||| || ||||  || | || | |  | |||| ||| | |||
Db    784 GAGAAGCCTGTCACCATCAGCGTGCATCAACCCCTGAAGGCTGCTCCTCCGCGGTGCAAGATGATC 843

Db    851 GAGAGGCCCATCAGCGTGCATCAACCCCTGAAGGCTGCTCCTCCGCGGTGCAAGATGATC 910

Db    844 CTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAAACTAGCCGAAGAGAGATTCCTCTGAAA 903

Qy   1919 CTGACTCAGGTAAAGCAGCACCAACAACAGAGGCTCTGCAAAGTGGACCAC 1970
          |||||||| |||||| | || || ||||||||||| ||||| || |||||
Db    911 TTGGAGATTATGCACAAGGAGCAAGTGAAAAACGGCACACCAAAAACCGGCAGAGTGAAGTTCCCCTGAAG 970

Db    904 ATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTGAAG 963

Db    971 ATCCTGGCTCATAACAACTTCGTCGGCGGACTCATTGGCAAGGAGGCCGGAACCTGAAG 1030

Qy    964 AAAATTGACATGACACAGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATA 1023

Db   1031 AAGGTGGAGCAGGAGCACACAGAGACGAAGATCACCATCACCATCATCGTCGGCTCCAGGACCTCACGCTC 1090

Qy   1024 TACACCCGAAAGACCAATCACCACTGTGAGGGCACCAGTTGAGGCCGTTGCCCAGTGCTGAG 1083

Db   1091 TATAACCCTGAGAGCGATCACTGTGAGGGCCCATTGAGAACTGTGTTGCAGGGCCGAG 1150

Qy   1084 ATAGAGATTATGAAGAAGCTGCGTGAGAGCTTTGAAAATGATATGCTGGCTGTAACCAA 1143

Db   1151 CAGGAGGATCATGAGAAGAAGTCGAGAGGTTACGGAGGCTTAGCGACGCGGCCGCCATGAGAGCTTG 1210

Qy   1144 CAAGCCAATCTGATCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGACTG 1203

Db   1211 CAGTCCCACCTCACCTCCTGGGCTTAACCTGGCTGGCTGTGAGGTCTTCTTCCCGCTGCCACCCCC 1270

Qy   1204 TCCGTGCTATCTCCACCAGCAGGGCCCTGGGCTGGACAGGTCCGGCTGAAGTCTACGCC 1263

Db   1271 AGGCGTGTCCCTCGTCCCTCCTCC-----TCCCAGCAGTGTCACTGGGGCTGCTCCTATAGCTCC 1324

Qy   1264 TTCGTGGACTACCCACCAGCAACTGGGCCATCGCGCGGATACCCCCGTCACCCCTCCGGGTAAA 1323

Db   1325 TTCA-------------------------------------- 1328

Qy   1324 CCGCATCATCACTCTTATCCAGAGCAGCAGATTGTGAATCTTCATCCAACCCAGAGC 1383

Db   1329 -------TGCAGGCTCCGGACGAGGAGATGGTACAAGTGTTCATCCCCGACCACAGGGCT 1378

Qy   1384 GTGGGGCCCATCATCGGGAAGAAGGGGCACACATCAAACAGCTGGCAGATTGGGATGACTT 1443
```

```
RESULT 14
US-09-261-855-1
; Sequence 1, Application US/09261855A
; Patent No. 6255055
; GENERAL INFORMATION:
;  APPLICANT: Ross, Jeffrey
;  TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN
;  TITLE OF INVENTION: (CRD-BP) AND ITS NUCLEIC ACID SEQUENCE
;  FILE REFERENCE: 960296.95131
;  CURRENT APPLICATION NUMBER: US/09/261,855A
;  CURRENT FILING DATE: 1999-03-03
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 1
;  LENGTH: 2224
;  TYPE: DNA
;  ORGANISM: Mus musculus
US-09-261-855-1

Query Match        20.3%;  Score 691.8;  DB 3;  Length 2224;
Best Local Similarity   63.5%;  Pred. No. 2.9e-183;
Matches   1152;  Conservative   0;  Mismatches  582;  Indels  81;  Gaps  3;

Qy     73 ATGAACAAGCTTACATCGGGAAAATCATGGAAGTCATGGAGTTCATTGAAAGTCTTAAAAAGCTAAGGAGC 132
          |||||||||||||||||||| ||  |||  | |||  ||||| || |||| | ||| | | || |||| |||
Db    131 ATGAACAAGCTTACATCGGCAACCTCAACGAGAGTGACCCCAGACTTGAGAAA 190

Qy    133 CTCTTTGGGACAGGAAGCTGCCCTGGCGGGACAGGTCCTGCTGGAAGTCTACGCC 192
          ||||| ||||  | |||| |||||| ||  |||||| | |||  ||||| ||||||
Db    191 GTATTCGCGAGCACAGATCTCCTACAGCGCCGCCCAGTTCTTGGTCAAATCGGCTGTACGCC 250

Qy    193 TTCGTGGACTACCCCGACCAGACACTCGGGCCATCGCGCGGATACCCCCTCCGGGTAAA 252
          |||||| |||||| | ||  | |||| ||| ||||| ||||| ||||| || |||| ||
Db    251 TTCGTGGATTGCCCCGACCAGACACTGGGCCATCGCGCGGATACCCCTCCGGGGAAA 310

Qy    253 GTGGAATTGCATGGGAAAATCATGCAGAAGTTGATTACTGAGTTCTTAAAAAGTAAGGAGC 312
          |||||||||| | | ||||  ||||| |||   ||||  |||||| |||||  |||||| |
Db    311 GTAGAACTGCAGAAGAAAAACGTCTAGAGATTGAACACTGAACACTCAGTCCCCAAAAAACAAAGGAGT 370

Qy    313 AGGAAAATTCAGATTCGAAACATCCCTCACCTGCAAACATCCCTCACCTGCAAAGGAGATTGGATGACTT 372
```

```
Db  1379  GTGGGGCCATCATTGGCAAGAAGGGCCCAGCACATCAAACAACTTCTCCCGTTTCGCCAGC  1438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   224  TGAGACGGCAGTGGTGATGTCACCTATTCCACCGGGAGCAGCAGCCAGGCAGCCATCAT   283

Qy  1444  GCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGTGTCATCATC  1503
          ||||  |||||||      ||||||| |||||||||||||||||||||| |||||  |||
Db   483  GAAGCTAAGCGGGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCGGATGA   542

Db  1439  GCCTCCATCAAGATTGCTCCTCCACCAGAAACACCTGACTCCAGGTCGTCGTCATC  1498
          ||||  ||||||||      |||    |||||  |||||||||||  |||||||||
Qy   284  GAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACATCCCGATGA   343

Qy  1504  ACCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAAGAG  1563
          |  |||              |    ||| ||||||||||||||||||||
Db   543  -------AGAGGTGAGCTCCCCTTCGCCCCCAGCGGCCAGCGTGGGGACCACTC        593

Db  1499  ACTGGACCCCCAGAGGCTCAGTTCAAGGCTCAGGGAGAATTTATGCGCAAACTAAAGAA  1558
          ||||||||||||      ||||| |||||     ||  ||||||||||||||
Qy   344  GCAGATAGCACAGGGACCTGAGAATGGGCGCGCAGGGGGCCCGGAGGGGGCTTTGGGGTCAGCC   403

Qy  1564  GAAAACTTCTTAACCCAAAGAAGAAGTGAAGCTGGAAGCGCATATCAGATGGTCCCTCT  1623
          |||||||||||||  |     || |    |||||||| ||||||  ||||||||||||
Db   594  TTCCCGGGAGCAAGGCCACGCCCCTGGGGCACTTCTCAGCGCCAGACAGATTGATTTCCC   653

Db  1559  GAGAATTTCTTGGTCCCAAGGAGGAAGTAGAAGCTAGAGACCCACCATACGGGTTCCGGCT  1618
          ||||  ||||  ||      |||| ||||||     |||      ||||||||||||||
Qy   404  CCGCAGGGCTCACCTGTGGCAGCGGCGGGGGCAGCGCAGCAGTGACAAGTGACATCCCC   463

Qy  1624  TCCAAGCTGCCGGGTGATTGGCAAGAGTGGCAAGACCGTGAACGAACTGCAGAACTTA  1683
          ||||||||||||||||||||||||||||   |||  ||||||||||||||||||||
Db   654  GCTGCGGATCCTCGTCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAGAGAACTC   713

Db  1619  TCAGCAGCCGCCGCGCTCATCGGCCAAGGCGGCAAAACCGTGAATGAGCTGCAGAACTTG  1678
          ||    ||  |||||    |||  |||  |||   |||||||||| |||||||||| ||
Qy   464  CCTTCGGCTCCTGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGAGGAGGAGGCCAC   523

Qy  1684  ACCAGTGCAGAAGTCATCGTGCCTCGTCGTGACCAAACGCCAGATGAAATGAGGAAGTGATC  1743
          ||||  |||||||||||  ||||||| |||||||||||    |||||| ||| ||||||
Db   714  CATAAAGAACATCACTAAGCAGCAGATCCATGCCACCCAGTAGATATCATAGAAAAGAACTC   773

Db  1679  ACTGCAGCTGAGGTGGTAGTGCCAAGAGACCAGAACCTGACCACCAAGTTCATT  1738
          ||    |||    ||| |||||    ||  |||    |||||||||||||||||
Qy   524  CATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGAGAGCGC   583

Qy  1744  GTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGGAAATT  1803
          ||  ||||     |||||||||      ||| |||||||||||||||||||||||||
Db   774  TGGAGCTGCAGAGAAGCCTGTCACCATCCATGCCACCCAGAGGGGACTTCTGAAGCATG   833

Db  1739  GTTAAGATCATCGGACATTTCTATGCCAGCCAGATGGCTCAGCCGAAGATCCGAGACATC  1798
          ||    ||||  |  ||||| || ||||  |||||   ||||| |||||||| ||||||
Qy   584  AGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCTGAGGGTCTGCTCCTCCGCTTG   643

Qy  1804  GTACAACAGGTGAAGCAGCAGGAGGAGAAATACCCTCAGGGAGTCGCCTTCACAGCGCAGC  1863
          ||||||||  ||| ||| ||| ||| |||   |||||||||||||||    |||||||||
Db   834  CCGGATGATTCTTGAAATCATGCAGAAGAGGCAGATGAGACCAAACTAGCCGAAGAGAT   893

Db  1799  CTGGCTCAAGTTAAGCAACAGCACCAGAGGGGACAGAGCAACCTGCCCCAGCCACGCGAGG  1858
          |||   || |||   ||||||||    |||    ||||||  ||    |||||||||||
Qy   644  TAAGATGATCTTGGAGATTATGCATAAAGAGCATAATAACTTTGAGGGGTCTCATTGGCAAGGAAGACG   703

Qy  1864  AAGTGAGGCTCCCAC  1878
          ||||||||||||||
Db   894  TCCTTCTGAAAAATCTTGGCACAACATGGCTTGGTTGGAAGGACTGATTGGAAAGAAGAAGCAG   953

Db  1859  AAGTGACCCCGCCCC  1873
          |||  |||  |||||
Qy   704  TCCCCTGAAGATCCTGGCCCATAATAACTTTGAGGGGTCTCATTGGCATTGGCAAGGAAGACG   763
```

```
Qy   954  AAATTTGAAGAAAATTGAACATGAAACATGAAACAGGACCAAGATAACAGATAACAATCTCATCCC  1013
          |||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db   764  GAACCTGAAGAGGTAGAGCAAGATACCGAGACAAAATCACCATCTCCTCGTTGCAAGA   823

Qy  1014  TTTGAGCATATACAACCCGGAAAGAACCATCACTGTGAAGGCACAGTTGAGGCCTGTGC  1073
          ||||||||||||||||||||||||||||||||||||||||||||   ||||||||||
Db   824  CCTTACCCTTTACAACCTGAGGAGACCATCATGGGGGCCATGGAGGGCGAAAGAATTGTTG   883

Db  1074  CAGTGCTGAGATAGAGATTATGAAGAGAGCTGCGTGAGGCCTTTGAAAAATGTATGTGGC  1133
          ||||  |||  || ||||  |||||  |||||||  |||||||||||||||||||||||
Qy   884  CAGGGCCCGAGCCAGGCAGGAAAATAATGAGAAAGTTCGGGAGGCCCATGAGAATGAGGATGGTGGCTGC   943
```

```
Qy  1134  TGTTAACCAACAGCCAATCGTCGTATCTGATCCGCAGGGTTGAACCTCAGCGCACTCGGCCATCTTTTC  1193
          |||||||||      |  |||||  ||||||||||||  ||||||||||||  ||||
Db   944  CATGAGCTCTCAC------CTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCC   997

Qy  1194  AACAGGACTGTCCGTGCTATCTCCACCAGAGGCCCCGGAGCTCCCCCGCTCCCCC  1253
          ||||||||||||||  |||||||||||||||      ||||||||||||||||
Db   998  AGCTTCATCCAGCGCAGTCCGCCGCCT------CCCAGCAGCGTTACTGGGGCTGCTCC   1051
```

```
Qy  1254  CTACCACCCCTTCACTACCCACTTCTCCGAGATACTTCTCCACCTGTACCCCATCACCAGTT  1313
          ||||||  ||||  ||||||   |||||||
Db  1052  CTATAGCTCCTTTATGC------------------------------------------   1068
```

```
Qy  1314  TGGCCCGTTCCCGACTCATCACTCTTATCCAGAGCAGGATTGTGAATCTCTTCATCCC  1373
          ||||||||||||||||||||||||||||
Db  1069  ------------------AGGCTCCCGACGGAGAGATGGAGACGAGGTTGTCAGGGTGTTTTATCCC   1105

Qy  1374  AACCCAGGCTGTGGCCGCCATCATCGGGAAGAAGGGGGCACATCAAACAGCTGGCGAG  1433
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db  1106  CGCCCAGGCAGTGGCCGGCGGCCATCATCGGCCAAGCATCATCAAACAGCTCTCCCG   1165

Qy  1434  ATTCGCCGGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGAAGCTCAGCGAAAGGAT  1493
          |||  |||||||||    ||||||||    |||| |||| ||||| ||||||||||||||
Db  1166  GTTTGCCAGCGCCGCTTCAACCACCGGAAACACCTGACTCCAAAGTTCGTAT   1225

Qy  1494  GGTCATCACCGGGCCCAGTTCAAGGCCAAGGGACGGATCTTTCGGAA  1553
          ||||||||||||||||||||||||||||||||   |||||||||| ||
Db  1494  GGTCATCACCGGGCCCAGTTCAAGGCCAAGGGACGGATCTTTCGGAA
```

RESULT 15
US-09-061-709-7
; Sequence 7, Application US/09061709B
Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 7
; LENGTH: 1946
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-7

Query Match   16.7%; Score 570.6; DB 3; Length 1946;
Best Local Similarity 63.0%; Pred. No. 2.5e-149;
Matches 997; Conservative 0; Mismatches 499; Indels 87; Gaps 4;

```
Qy   303  GCTAAGGACGAGGAGAAATTCAGATTCGAAACATCCCTCCACCTGCAGTCGGGAGGTGTT  362
          |||  |||||| || || |||  ||||||||||  |||   |||||||||||||||||
Db   104  GGTTAGGAGCCGAAAAATTCAAATCCGAAATATTCCACCCCAGTCCGAATGGAAGTACT   163

Qy   363  GGATGGACTTTGGCTCAATATGGCTCAATATGGCTCAGTAGTGTGAGAATGTGAACAAGTCAACAGACAC  422
          ||||    |||||||   |||||  ||||  ||||||||||||  |||||||||||||||||
Db   164  GGACAGCGCCTGCTGTGGCTGCATGGTGACGTACAGTAGGAACTGTGAGCAAGTGAACACCCGAGAG   223

Qy   423  AGAAACCGCCGTTGCAACGTTCACATATGCAACAAGAAGAAGAAGAAAATAGCCATGA  482
          || || ||| ||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  1226 GGTTATCATCACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAA 1285

Qy  1554 ACTGAAAGAGGAAAACTTCTTTAACCCCAAAGAAGAGAAGTGAAGCTGGAAGCGCATATCAG 1613
          ||| ||||||||  ||||| |||||    ||    || ||||||||||  || | ||  |  |
Db  1286 ACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGACACCCACATACG 1345

Qy  1614 AGTGCCCTCTTCCACAGCTGCCGGGTGATTGGCAAAGTGGCAAGACCGTGAACGAACT 1673
          ||||| | |  | |   |||||||||||| || |||||| | ||| | |||||||| |
Db  1346 TGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTT 1405

Qy  1674 GCAGAACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGA 1733
          ||||||||  ||||  | |||  ||     |  |||     ||||| || |||| || ||
Db  1406 GCAGAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACACCAGACCCCTGATGAGAACGA 1465

Qy  1734 GGAAGTGATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGCTGCACAGCGCCAAGAT 1793
          |||||| | |  |||| |||| ||| || || |  |  ||| || | |||| |||||||
Db  1466 CCAGGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGAT 1525

Qy  1794 CAGGGAAAATTGTACAACAGTGAAGCAGCAGCAGAAATACCCTCAGGGAGTCGCCTC 1853
          |||||  | |||| ||||    |||||||| |||     |||||  ||  | ||
Db  1526 CCGAGACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCCA 1585

Qy  1854 ACAGCCGCAGCAAGTGAGGCTCCC 1876
          |||| || |||||| ||| |||
Db  1586 GGCACGGAGGAAGTGACCAGCCC 1608
```

Search completed: July 21, 2004, 12:27:13
Job time : 255 secs

```
          GenCore version 5.1.6
     Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    July 21, 2004, 03:46:00 ; Search time 1509 seconds
                                      (without alignments)
                              11031.060 Million cell updates/sec

Title:           US-09-270-437D-6
Perfect score:   3412
Sequence:        1 ggcagcggaggaggcgagga........aaccttgaaaatgtttattt 3412

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        3191023 seqs, 2439312756 residues

Total number of hits satisfying chosen parameters:     6382046

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  Published_Applications NA:*
      1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
      2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
      3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
      4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
      5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
      6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
      7: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
      8: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
      9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
     10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
     11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
     12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
     13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
     14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
     15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
     16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
     17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
     18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
     19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3411	100.0	3412	9	US-09-899-651-6	Sequence 6, Appli
2	3320.6	97.3	3694	9	US-09-764-864-329	Sequence 329, App
3	3143	92.1	3283	9	US-09-889-651-8	Sequence 8, Appli
4	3116.4	91.3	3905	10	US-09-814-353-20241	Sequence 20241, A
5	3114.8	91.3	3667	15	US-10-097-340-146	Sequence 146, App
6	3114.8	91.3	3667	17	US-10-648-593-48	Sequence 48, Appl
7	1989.8	58.3	2290	9	US-09-764-864-332	Sequence 332, App
8	1423.8	41.7	1707	16	US-10-262-445-13	Sequence 39, Appl
9	1096.8	32.1	1186	9	US-09-764-864-749	Sequence 749, App
10	776.2	22.7	822	9	US-09-764-864-330	Sequence 330, App
11	712	20.9	1740	9	US-09-735-705-347	Sequence 347, App
12	712	20.9	1740	9	US-09-850-716A-347	Sequence 347, App
13	712	20.9	1740	9	US-09-897-778-347	Sequence 347, App
14	712	20.9	1740	13	US-10-007-700-347	Sequence 347, App
15	712	20.9	1740	15	US-10-117-982-347	Sequence 347, App
16	712	20.9	1740	15	US-10-117-982-478	Sequence 478, App
17	712	20.9	1740	16	US-10-313-986-347	Sequence 347, App
18	712	20.9	1740	16	US-10-313-986-478	Sequence 478, App
19	712	20.9	1743	9	US-09-897-778-447	Sequence 447, App
20	712	20.9	1743	9	US-09-897-778-450	Sequence 450, App
21	712	20.9	1743	13	US-10-007-700-447	Sequence 447, App
22	712	20.9	1743	13	US-10-007-700-450	Sequence 450, App
23	712	20.9	1743	15	US-10-117-982-447	Sequence 447, App
24	712	20.9	1743	15	US-10-117-982-450	Sequence 450, App
25	712	20.9	1743	16	US-10-313-986-447	Sequence 447, App
26	712	20.9	1743	16	US-10-313-986-450	Sequence 450, App
27	712	20.9	1799	16	US-10-313-986-485	Sequence 485, App
28	712	20.9	34555	15	US-10-117-982-479	Sequence 479, App
29	712	20.9	34555	16	US-10-313-986-479	Sequence 479, App
30	711.8	20.8	1740	15	US-10-117-982-483	Sequence 483, App
31	711.8	20.8	1740	16	US-10-313-986-483	Sequence 483, App
32	710.4	20.8	4159	9	US-09-899-651-4	Sequence 4, Appli
33	710.4	20.8	4181	9	US-09-735-705-175	Sequence 175, App
34	710.4	20.8	4181	9	US-09-954-456-715	Sequence 715, App
35	710.4	20.8	4181	9	US-09-850-716A-175	Sequence 175, App
36	710.4	20.8	4181	9	US-09-897-778-175	Sequence 175, App
37	710.4	20.8	4181	10	US-09-466-396A-175	Sequence 175, App
38	710.4	20.8	4181	13	US-10-007-700-175	Sequence 175, App
39	710.4	20.8	4181	15	US-10-117-982-175	Sequence 175, App
40	710.4	20.8	4181	16	US-10-313-986-175	Sequence 175, App
41	710.4	20.8	4434	16	US-10-116-802-145	Sequence 145, App
42	709	20.8	1764	13	US-09-850-716A-428	Sequence 428, App
43	709	20.8	1764	9	US-09-897-778-428	Sequence 428, App
44	709	20.8	1764	13	US-10-007-700-428	Sequence 428, App
45	709	20.8	1764	15	US-10-117-982-428	Sequence 428, App

ALIGNMENTS

```
RESULT 1
US-09-899-651-6
; Sequence 6, Application US/09899651
; Patent No. US2002011470A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Eike
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
; TITLE OF INVENTION: Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/899,651
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/09/061,709
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 6
; LENGTH: 3412
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-899-651-6

Query Match             100.0%;  Score 3411;  DB 9;  Length 3412;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 3412;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy   1 GGCAGCGGAGGAGGCGAGGAGCGAGGAGCGGCGCCGGGTACCGGCGGGGGGAGCGGGGTCTTCGGGG 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGCAGCGGAGGAGGCGAGGAGCGAGGAGCGGCGCCGGGTACCGGCGGGGGGAGCGGGGTCTTCGGGG 60

Qy  61 AAGAGACGGGATGATGAACAAGCTTTACATCGGGAACTGGAGCCCCGGTCCCGCGCCGGCCGGAC 120
```

DNA sequence alignment (Qy / Db) spanning positions 61–2280.

Qy   3361 AGCCAAGAACCNATATGCCTTCTTCTTTTGGACAAACCTTGAAAAATGTTTATTT 3412
            |||||||||| |||||||||||||||||| |||||||||||||||||||||||||
Db   3361 AGCCAAGAACCNATATGCCTTCTTCTTTTTGGACAAAACCTTGAAAATGTTTATTT 3412

RESULT 2
US-09-764-864-329
; Sequence 329, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;   FILE REFERENCE: PTZ23
;   CURRENT APPLICATION NUMBER: US/09/764,864
;   CURRENT FILING DATE: 2001-01-17
;   Prior application data removed - consult PALM or file wrapper
;   NUMBER OF SEQ ID NOS: 1792
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 329
;   LENGTH: 3694
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-864-329

Query Match   97.3%; Score 3320.6; DB 9; Length 3694;
Best Local Similarity   99.9%; Pred. No. 0;
Matches 3343; Conservative 1; Mismatches 1; Indels 3; Gaps 2;

Qy   68 GGATGATGAACAAGCTTTACATCGGAACCTGAGCCCCGCGTCACCGCCGACGACCTCC   127
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   91 GGATGATGAACAAGCTTTACATCGGAACCTGAGCCCCGCGTCACCGCCGACGACCTCC   150

Qy   128 GGCAGCTCTTTGGGGACAGGAGAGCTGCCCCTGGCGGGGACAGGTGCCCCTGGCGGGGACAGCA   187
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   151 GGCAGCTCTTTGGGGACAGGAGAGCTGCCCCTGGCGGGGACAGGTGCCCCTGGCGGGGACAGCA   210

Qy   188 ACGCCTTCGTGACTACCCGACCAGAACTGGGAGAATGTGGAACAAGTCAACACAGAAA   247
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   211 ACGCCTTCGTGACTACCCGACCAGAACTGGGAGAATGTGGAACAAGTCAACACAGAAA   270

Qy   248 GTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAGCTAA   307
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   271 GTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAGCTAA   330

Qy   308 GGAGCAGGAAAATTCAGATTCGAAACATCCCTCCTCAGTGGAGGTGTTGGATG   367
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   331 GGAGCAGGAAAATTCAGATTCGAAACATCCCTCCTCAGTGGAGGTGTTGGATG   390

Qy   368 GACTTTTGGCTCAATATGGACAGTGGAGAATGTGGAACAAGTCAACAGAACAGAAA   427
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   391 GACTTTTGGCTCAATATGGACAGTGGAGAATGTGGAACAAGTCAACAGAACAGAAA   450

Qy   428 CCGCCGTTGTCAAGTCACATCACATATGCAACAAGAGAGAAGCAAAAATAGCCATGGAGAAGC   487
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   451 CCGCCGTTGTCAAGTCACATCACATATGCAACAAGAGAGAAGCAAAAATAGCCATGGAGAAGC   510

Qy   488 TAAGCGGCATCAGTTTGAGACTACTCCTTCAAGATTTCCTACATCCCGGATGAAGAGAG   547
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   511 TAAGCGGCATCAGTTTGAGACTACTCCTTCAAGATTTCCTACATCCCGGATGAAGAGAG   570

Qy   548 TGAGCTCCCCTTCGCCCCCTCAGCGGACCGTGGGAGCCAG   607
           ||||||||||||||||||||||||||||||||||||||||
Db   571 TGAGCTCCCCTTCGCCCCCTCAGCGGACCGTGGGAGCCAG   630

Qy   608 GCCACCCAGTTCGGGGGCACTTCTCAGGCCAGACAGATTGATTCCCGCTCGCGGATCCTGG   667
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   631 GCCACCCAGTTCGGGGGCACTTCTCAGGCCAGACAGATTGATTCCCGCTCGCGGATCCTGG   690

Qy   668 TCCCACCCAGTTGTTGGTGCCATCATCCGGAAAGGAGGGGCTTGACCATAAGAACATCA   727
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   691 TCCCACCCAGTTGTTGGTGCCATCATCCGGAAAGGAGGGGCTTGACCATAAGAACATCA   750

Qy   728 CTAAGCAGACACCCAGTCCCGGGTAGATATCCGGGTAGATATCGGAGCTGCGAGAGA   787

Qy   2281 GGGAGAGATGTTAAGATATGTGGCCTGTGGTTACACAGGGTGCCTGCGCGGTAATATAT   2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 GGGAGAGATGTTAAGATATGTGGCCTGTGGTTACACAGGGTGCCTGCGCGGTAATATAT   2340

Qy   2341 TTTAGAAATAATATATCAAATAACTCAATTTTTAATCAATTATTAATTAATTTT   2400
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 TTTAGAAATAATATATCAAATAACTCAATTTTTAATCAATTATTAATTAATTTT   2400

Qy   2401 TTTTTCTTTTAAAGAGAAGCAGCAGCGCTTTCTAGACTTTTAAAGAATAAAGTCTTGGGAG   2460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 TTTTTCTTTTAAAGAGAAGCAGCAGCGCTTTCTAGACTTTTAAAGAATAAAGTCTTGGGAG   2460

Qy   2461 GTCTCACGGTGTGAGAGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAAGGGATA   2520
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 GTCTCACGGTGTGAGAGGACACTCACGGCAGTTCTGAGGCCACCCGCACAAAAATTCACCCAGACGCGAAAT   2520

Qy   2521 CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAAGGGATA   2580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAAGGGATA   2580

Qy   2581 CCGTCTCCTGAAGAGGAAACTCTGCCACTCCTCAGCGCGTCTAGCTCATACACCCATT   2640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 CCGTCTCCTGAAGAGGAAACTCTGCCACTCCTCAGCGCGTCTAGCTCATACACCCATT   2640

Qy   2641 TCTCTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATAATAATTCCTGTCT   2700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 TCTCTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATAATAATTCCTGTCT   2700

Qy   2701 CTCTCGTTTATCTCTCCCCTTCCCCCCTTCTCTCCATCTCCATTCTTCTTTGAA   2760
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 CTCTCGTTTATCTCTCCCCTTCCCCCCTTCTCTCCATCTCCATTCTTCTTTGAA   2760

Qy   2761 TTTCCTCATCCTCCATCTCAATCCGTATCTACGGCACCCCCCCCCCCCCCCAGGCAAAGCA   2820
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 TTTCCTCATCCTCCATCTCAATCCGTATCTACGGCACCCCCCCCCCCCCCCCAGGCAAAGCA   2820

Qy   2821 GTGCTCTGAGTATCACATCACACAAAGAACAAAAAGCGAAACACACAGAAACCAGCCTCAA   2880
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 GTGCTCTGAGTATCACATCACACAAAGAACAAAAAGGCAACAAAGCACACACAGAAACCAGCCTCAA   2880

Qy   2881 CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGGGTTTTGGAAG   2940
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGGGTTTTGGAAG   2940

Qy   2941 AGGAAAACAGGAACCCACCACCAACCAATCAACCAAACAAGAAAAAATTCCACAATG   3000
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 AGGAAAACAGGAACCCACCACCAACCAATCAACCAAACAAGAAAAAATTCCACAATG   3000

Qy   3001 AAAGAATGTATTTTGTCTTTTTGTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATG   3060
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 AAAGAATGTATTTTGTCTTTTTGTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATG   3060

Qy   3061 ATTCCCTTTCTTTAAAAAAAAATGTGGAGAAAGTAGAAATTTACCAAGGTTGTTGGCC   3120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 ATTCCCTTTCTTTAAAAAAAAATGTGGAGAAAGTAGAAATTTACCAAGGTTGTTGGCC   3120

Qy   3121 CAGGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAACAACAGAAGAAGCTACCTCAG   3180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 CAGGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAACAACAGAAGAAGCTACCTCAG   3180

Qy   3181 GTGTTTTTACCCTGAGAAGGTGACAGTCCAAGTGTGCAACAGCTTGTCAAGAGATATCAACT   3240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 GTGTTTTTACCCTGAGAAGGTGACAGTCCAAGTGTGCAACAGCTTGTCAAGAGATATCAACT   3240

Qy   3241 TTGGAGCATTTTTTATTTTTTTATAAAAATGAGTTGGAAAAAAATAAGATATCAACT   3300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 TTGGAGCATTTTTTATTTTTTTATAAAAATGAGTTGGAAAAAAATAAGATATCAACT   3300

Qy   3301 GCCAGCCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTCTGACCATAAAGAACATCA   3360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 GCCAGCCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTCTGACCATAAAGAACATCA   3360

Db   1831 AACAGGTGAAGCAGGAGCAGGAGCAGAAATACCCTCAGGGAGTCGGCCTCACAGCGCAGCAGCAAGT 1890
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1868 GAGGCTCCCACAGGCACCACCAGCAAAACAACACGATGAATGAATGTAGCCCTTCCAACACCTGACAG 1927
Db   1891 GAGGCTCCCACAGGCACCACCAGCAAAACAACACGATGAATGAATGTAGCCCTTCCAACACCTGACAG 1950

Qy   1928 AATGAGACCAAACGCAGCCAGCCAGATGGGAGCAAACCAATCTGAGGAATGAG 1987
Db   1951 AATGAGACCAAACGCAGCCAGCCAGATGGGAGCAAACCAATCTGAGGAATGAG 2010

Qy   1988 AAGTCTGCGGAGGCGGCCAGGAGACTCGCCGAGCCCTGAGAACCCGAGGAG 2047
Db   2011 AAGTCTGCGGAGGCGGCCAGGAGACTCGCCGAGCCCTGAGAACCCGAGGAG 2070

Qy   2048 GGGCGGGGAAGGTCAGCCAGGTTTGCCAGAACCACCGCAGTTCGGCCGCCCCAGG 2107
Db   2071 GGGCGGGGAAGGTCAGCCAGGTTTGCCAGAACCACCGCAGTTCGGCCGCCCCAGG 2130

Qy   2108 GCTTCTGCAGGCTTCAGCCATCCACTCACCGGATCTCCTCCTGAACTCCCAC 2167
Db   2131 GCTTCTGCAGGCTTCAGCCATCCACTCACCGGATCTCCTCCTGAACTCCCAC 2190

Qy   2168 GACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAGCCAAGCAAAATGC 2227
Db   2191 GACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAGCCAAGCAAAATGC 2250

Qy   2228 ACACCCTTTTTCTGTGTGGCAAATCGTCTCTGTACATGTGTACATTAGAAGGGAAGA 2287
Db   2251 ACACCCTTTTTCTGTGTGGCAAATCGTCTCTGTACATGTGTACATTAGAAGGGAAGA 2310

Qy   2288 TGTTTAAGATATGTGGCCTGTGGTTACACAGGGTGCTCAGCGCGGTAATATATTTTAGAA 2347
Db   2311 TGTTTAAGATATGTGGCCTGTGGTTACACAGGGTGCTCAGCGCGGTAATATATTTTAGAA 2370

Qy   2348 ATAATATATCAAATAACTCAACTAACTCCAATTTTTAATCAATTATTAATTTTTTCT 2407
Db   2371 ATAATATATCAAATAACTCAACTAACTCCAATTTTTAATCAATTATTAATTTTTTCT 2430

Qy   2408 TTTTAAAGAGAAAGCAGGCTTTCTCAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTCAC 2467
Db   2431 TTTTAAAGAGAAAGCAGGCTTTCTCAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTCAC 2490

Qy   2468 GGTGTAGAGAGGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGAGGGAAATCTCGTCG 2527
Db   2491 GGTGTAGAGAGGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGAGGGAAATCTCGTCG 2550

Qy   2528 GAAGGACACTCACGGACAGTTCTGGATCACCGTGTATGTCAACAGAAGGGATACCGTCTC 2587
Db   2551 GAAGGACACTCACGGACAGTTCTGGATCACCGTGTATGTCAACAGAAGGGATACCGTCTC 2610

Qy   2588 CTTGAAGAGGAAACTCGTCACTCTCCTCATGCCTCTAGCTGCTATACACCCATTTCTCTT 2647
Db   2611 CTTGAAGAGGAAACTCGTCACTCTCCTCATGCCTCTAGCTGCTATACACCCATTTCTCTT 2670

Qy   2648 GCTTCACAGGTTTAAACTGGTTTTTGCATACTGGTATACTGCTATATAATTCTCTGTCTCTCTG 2707
Db   2671 GCTTCACAGGTTTAAACTGGTTTTTGCATACTGGTATACTGCTATATAATTCTCTGTCTCTCTG 2730

Qy   2708 TTTATCTCTCCCCATCTCAATCCCGTATCTCACGCACCCCCCCCCCGTATCTCATCTTTTGAATTTCCTC 2767
Db   2731 TTTATCTCTCCCCATCTCAATCCCGTATCTCACGCACCCCCCCCCCGTATCTCATCTTTTGAATTTCCTC 2790

Qy   2768 ATCCCTCCAATCCCGTATCTCACGCA--CCCCCCCCCCCCCAGGCAAAGCAGTGCT 2825
Db   2791 ATCCCTCCAATCCCGTATCTCACGCACCCCCCCCCCCCCCCAGGCAAAGCAGTGCT 2850

Qy   2826 CTGAGTATCACATCACACACAAGAACAAAAGCAAACACAAACCAGCCAACTTAC 2885
Db   2851 CTGAGTATCACATCACACACAAGAACAAAAGCAAACACAAACCAGCCAACTTAC 2910

Qy   2886 ACTTGGTTACTCACAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGGAA 2945
Db   2911 ACTTGGTTACTCACAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGGAA 2970

Db   751 |||||||||||||
          CTTAAGCAGACCCAGTCCCGGGTAGATATCCAAGAGAAAGAGAACTTCGGAGCTGCAGAGA 810
Qy   788 AGCCTGTCACCATCCATGCCACCCCCAGAGGGGACTTCTGAAGCATGCCGCATGATTCTTG 847
Db   811 AGCCTGTCACCATCCATGCCACCCCCAGAGGGGACTTCTGAAGCATGCCGCATGATTCTTG 870

Qy   848 AAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAGAGATTCCTCTGAAAATCT 907
Db   871 AAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAGAGATTCCTCTGAAAATCT 930

Qy   908 TGGCACACAAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTGAAGAAA 967
Db   931 TGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTGAAGAAA 990

Qy   968 TTGAAACATGAAACAGGACCAAGAGATAACAATCTCATCTTTGCAGGATTTGAGCATATACA 1027
Db   991 TTGAAACATGAAACAGGACCAAGAGATAACAATCTCATCTTTGCAGGATTTGAGCATATACA 1050

Qy   1028 ACCCGAAAGAACCATCACTGTGAAGGCACAGTTGAGGCCTGTCGTGCCAGTGCTGAGATAG 1087
Db   1051 ACCCGAAAGAACCATCACTGTGAAGGCACAGTTGAGGCCTGTCGTGCCAGTGCTGAGATAG 1110

Qy   1088 AGATTATGAAGAAGTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTAACCAACAG 1147
Db   1111 AGATTATGAAGAAGTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTAACCAACAG 1170

Qy   1148 CCAATCTGATCCCAGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGACTGTCCG 1207
Db   1171 CCAATCTGATCCCAGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGACTGTCCG 1230

Qy   1208 TGCTATCTCCACCAGCAGGAGCTCCGCGGAGCTCCCCGCTGCCCCGTACCACCCCCTTCA 1267
Db   1231 TGCTATCTCCACCAGCAGGAGCTCCGCGGAGCTCCCCGCTGCCCCGTACCACCCCCTTCA 1290

Qy   1268 CTACCCACTCCGGATACTTCTCCAGCCTGTACCCGTTGGCCCGGTTCCCGC 1327
Db   1291 CTACCCACTCCGGATACTTCTCCAGCCTGTACCCGTTGGCCCGGTTCCCGC 1350

Qy   1328 ATCATCACTCTATCGGAAGCAGGAGATTGTGAATCTCTTCATCCGAACCAGGCTGTGG 1387
Db   1351 ATCATCACTCTATCGGAAGCAGGAGATTGTGAATCTCTTCATCCGAACCAGGCTGTGG 1410

Qy   1388 GCGGCCATCATCGGGAAGACGGGGCCACATCAAACAGCTGGCGAGATTCGCCGGAGCCT 1447
Db   1411 GCGGCCATCATCGGGAAGACGGGGCCACATCAAACAGCTGGCGAGATTCGCCGGAGCCT 1470

Qy   1448 CTATCAAGAGATTGCCCGGAAGCCCAGTTCAGGCCCAGGAAAGGATGGTCATCATCACCG 1507
Db   1471 CTATCAAGAGATTGCCCGGAAGCCCAGTTCAGGCCCAGGAAAGGATGGTCATCATCACCG 1530

Qy   1508 GGCCACCGGAAGCCCAGTTCAGGCCCAGTCCTGCAAGACCGGATCTTTGGGAAACTGAAAGAGGAAA 1567
Db   1531 GGCCACCGGAAGCCCAGTTCAGGCCCAGTCCTGCAAGACCGGATCTTTGGGAAACTGAAAGAGGAAA 1590

Qy   1568 ACTTCTTTAACCCAAAGAAGAGAGTGAAGCTGGAAGCCATATCAGAGTGCCCTCTTCCA 1627
Db   1591 ACTTCTTTAACCCAAAGAAGAGAGTGAAGCTGGAAGCCATATCAGAGTGCCCTCTTCCA 1650

Qy   1628 CAGCTGGCCGGGTGATTGGCAAAGGTGGCAAGAACCGTGAACGCAAGATCGCAGAACTTAACCA 1687
Db   1651 CAGCTGGCCGGGTGATTGGCAAAGGTGGCAAGAACCGTGAACGCAAGATCGCAGAACTTAACCA 1710

Qy   1688 GTTGCAGAAGTCATCGGGCACTTCTTTGCTCGTGCTCGCAGACTTGCACAGGCGCAAGATCGCAGAAATTGTAC 1747
Db   1711 GTTGCAGAAGTCATCGGGCACTTCTTTGCTCGTGCTCGCAGACTTGCACAGGCGCAAGATCGCAGAAATTGTAC 1770

Qy   1748 GAATTATCGGGCACTTCTTTGCTCGTGCTCGCAGACTTGCACAGGGAGCTCGCAGACTTGTAC 1807
Db   1771 GAATTATCGGGCACTTCTTTGCTCGTGCTCGCAGACTTGCACAGGGAGCTCGCAGAAATTGTAC 1830

Qy   1808 AACAGGTGAAGCAGCAGGAGCAGGAGCAGAAATACCCTCAGGGAGTCGGCCTCACAGCGCAGCCAGT 1867

```
Qy   121 GACTTCCGGCAGCTCTTTGGGACCAGCAGAAGCTGCCCTGCGGGGACCAGAGGTCTGCTGCTGAAG 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GACTTCCGGCAGCTCTTTGGGACCAGCAGAAGCTGCCCTGCGGGGACCAGAGGTCTGCTGCTGAAG 180

Qy   181 TCCGGCTACGCCTTCGTGACTACCACCGACCCAGCAGGACTCGGCGCATCGGCGCCATCGCGAGACC 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TCCGGCTACGCCTTCGTGACTACCACCGACCCAGCAGGACTCGGCGCATCGGCGCCATCGCGAGACC 240

Qy   241 CTCTCGGGTAAGTGCATGGGAAATGCATGGGAAATCATGGGAAGTTGATTACTCAGTCTCTAAA 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CTCTCGGGTAAGTGCATGGGAAATGCATGGGAAATCATGGGAAGTTGATTACTCAGTCTCTAAA 300

Qy   301 AAGCTAAGGAGCAGGAAAATTCAGATTCGAAACACCCTCCTCACCTGCAGTGGGAGGTGG 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AAGCTAAGGAGCAGGAAAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGGAGGTGT 360

Qy   361 TTGGATGGAGCTTTTGGCTCAATATGGGACAGTGGGAACAGAGCAGTCAACACAGAC 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TTGGATGGAGCTTTTGGCTCAATATGGGACAGTGGGAACAGAGCAGTCAACACAGAC 420

Qy   421 ACAGAAACCGCCGTTGCACGTCACATATGCAACAGAGAAGCAAAAATAAGCCATG 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ACAGAAACCGCCGTTGCACGTCACATATGCAACAGAGAAGCAAAAATAAGCCATG 480

Qy   481 GAGAGCTAAGCGGCATCAGTTGAGAACTACTCCTTCAAGATTTCTTCCGCTAGCC 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GAGAGCTAAGCGGCATCAGTTGAGAACTACTCCTTCAAGATTTCTTCCGCTAGCC 540

Qy   541 GAAGAGGTGAGCTCCCCTTCGCCCCTGGGGGCCACTCTCAGGCCAGACAGATTGATTTCCCGCTGCGG 600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GAAGAGGTGAGCTCCCCTTCGCCCCTGGGGGCCACTCTCAGGCCAGACAGATTGATTTCCCGCTGCGG 600

Qy   601 GAGCAGGAAGGCCACGCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGG 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GAGCAGGAAGGCCACGCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGG 660

Qy   661 ATCCTGGTCCCCACCAGTTGTTGGTGCCATCATCGGAAGAGGGCTTGACCATAAAG 720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 ATCCTGGTCCCCACCAGTTGTTGGTGCCATCATCGGAAGAGGGCTTGACCATAAAG 720

Qy   721 AACATCACTAAGCAGACCCAGTCCGGGTAGATATCCATAGAAAGAGAACTCTGGAGCT 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 AACATCACTAAGCAGACCCAGTCCGGGTAGATATCCATAGAAAGAGAACTCTGGAGCT 780

Qy   781 GCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGACTTCTGAAGCATGCCGCATG 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGACTTCTGAAGCATGCCGCATG 840

Qy   841 ATTCTTGAAATCATGCAGAAAGAGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 ATTCTTGAAATCATGCAGAAAGAGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG 900

Qy   901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAGAAGAAGGCAGAAATTTG 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAGAAGAAGGCAGAAATTTG 960

Qy   961 AAGAAAATTGAACATGAAACAGGACCAAGATAACAATCTCATCTTTGAGGCCTTGCCAGTGCT 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 AAGAAAATTGAACATGAAACAGGACCAAGATAACAATCTCATCTTTGAGGCCTTGCCAGTGCT 1020

Qy  1021 ATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAAGGGCCTTGCCAGTGCT 1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 ATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAAGGGCCTTGCCAGTGCT 1080

Qy  1081 GAGATAGAGATTATGAAGAAGCTGCGTGAAAATGATATGTGGCTGTTAAC 1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 GAGATAGAGATTATGAAGAAGCTGCGTGAAAATGATATGTGGCTGTTAAC 1140

Qy  1141 CAACAAGCAATCTGATCCCAGGGTTGACCTTCAGCGCGCACTTGACCGCATCTTTTCAACAGGA 1200
Db  1141 --------------------------------------------------------------- 1140
```

```
Qy  2946 AACAGGAACCCACCCAACCACCAATCAACCAACCAAACAAGAAAAATTCCACAATGAAAGA 3005
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2971 AACAGGAACCCACCCAACCACCAATCAACCAACCAAACAAGAAAAATTCCACAATGAAAGA 3030

Qy  3006 ATGTATTTGTCTTTTTGCATTTGGTGTATAAGCCATCCATCAATATTCAGCAAAATGATTCC 3065
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3031 ATGTATTTGTCTTTTTGCATTTGGTGTATAAGCCATCCATCAATATTCAGCAAAATGATTCC 3090

Qy  3066 TTTCTTT-AAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCCCAGG 3124
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3091 TTTCTTTAAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCCCAGG 3150

Qy  3125 GCGTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAAGAAGCTACCTCAGGTGT 3184
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3151 GCGTTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAAGAAGCTACCTCAGGTGT 3210

Qy  3185 TTTTACCTCAGCACCTTGCTCTTGTTGTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGG 3244
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3211 TTTTACCTCAGCACCTTGCTCTTGTTGTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGG 3270

Qy  3245 AGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACTGCCA 3304
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3271 AGCATTTTTTATTTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACTGCCA 3330

Qy  3305 GCCTGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCGAATTGTCTTCCGCTAGCC 3364
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3331 GCCTCGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCGAACTACTCCTTCGCTAGCC 3390

Qy  3365 AAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAAATGTTTATTT 3412
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  3391 AAGAACCTATATGGGCCTTCTTTTGGACAAACCTTGAAAAATGTTTATTT 3438
```

RESULT 3
US-09-899-651-8

; Sequence 8, Application US/09899651
; Patent No. US2002011470A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
; TITLE OF INVENTION: Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/899,651
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/09/061,709
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 8
; LENGTH: 3283
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-899-651-8

Query Match   92.1%; Score 3143; DB 9; Length 3283;
Best Local Similarity   96.2%; Pred. No. 0;
Matches 3283; Conservative 0; Mismatches 0; Indels 129; Gaps 1;

```
Qy     1 GGCAGCCGGAGGAGGCGGAGGAGCCGGGTACCGGGGAGCCGGGCCGGGCCTCTCGGGG 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GGCAGCCGGAGGAGGCGGAGGAGCCGGGTACCGGGGAGCCGGGCCGGGCCTCTCGGGG 60

Qy    61 AAGAGACGGATGATGAACAAGCTTTACATCGGAACAACCTGAGCCCCCGCCGTCACCGCCGAC 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 AAGAGACGGATGATGAACAAGCTTTACATCGGAACAACCTGAGCCCCCGCCGTCACCGCCGAC 120
```

```
Qy   1201  CTGTCCGTGCTATCTCCACCAGCAGGCCCCGCGGAGCTCCCCGCTGCCCCCTACCAC  1260
           
Db   1141  -------------------------------------------------------  1140

Qy   1261  CCCTTCACTACCTACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTTGGCCCCG  1320
           ||||||||||||||||||
Db   1141  --------ACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTTGGCCCCG  1191

Qy   1321  TTCCCGCCATCATCACTCTATCCGAGACAGGAGATTGTGAATCTTCATCCCAACCCAG  1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1192  TTCCCGCCATCATCACTCTATCCGAGACAGGAGATTGTGAATCTTCATCCCAACCCAG  1251

Qy   1381  GCTGTGGGCGCCATCATCGGGAAGAAGGGGGCCACACATCAAACAGCTGCGAGATTCGCC  1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1252  GCTGTGGGCGCCATCATCGGGAAGAAGGGGGCCACACATCAAACAGCTGCGAGATTCGCC  1311

Qy   1441  GGAGCCCTTATCAAGATTGCCCCTGCGGAAGCCCAGACGTCAGCGGAAGGATGGTCATC  1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1312  GGAGCCCTTATCAAGATTCAAGATTGCCCCTGCGGAAGCCCAGACGTCAGCGGAAGGATGGTCATC  1371

Qy   1501  ATCACCGGGCACCCGGAAGCCAGTTCAAGGCCCAGGGACACGGATCTTTGGGAAACTGAAA  1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1372  ATCACCGGGCACCCGGAAGCCAGTTCAAGGCCCAGGGACACGGATCTTTGGGAAACTGAAA  1431

Qy   1561  GAGGAAAACTTCTTTAACCCGAAAGAAGAGTGAAGCTGCAAGACCGTGAACGAACTCAGAC  1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1432  GAGGAAAACTTCTTTAACCCGAAAGAAGAGTGAAGCTGCAAGACCGTGAACGAACTCAGAC  1491

Qy   1621  TCTTCCACAGTGGCCGGGGTGATTGCAAAGGTGCAAGACCGTGAACGAACTCGAGAC  1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1492  TCTTCCACAGTGGCCGGGGTCATCGTGACCAAGATGAAAATGAGGAAGTG  1551

Qy   1681  TTAACCAGTGCAGAGTCATCGTGCCCTGTGACCAAACGCCAGATGAAAATGAGGAAGTG  1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1552  TTAACCAGTGCAGAGTCATCGTGCCCTGTGACCAAACGCCAGATGAAAATGAGGAAGTG  1611

Qy   1741  ATCGTCAGATTATCGGGCACTTCTTGCTTTGCTTGCTAGCCAAGATCGCAACGCGAAGA  1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1612  ATCGTCAGATTATCGGGCACTTCTTGCTTTGCTTGCTAGCCAAGATCGCAACGCGAAGA  1671

Qy   1801  ATTGTACACAGGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGC  1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1672  ATTGTACACAGGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGC  1731

Qy   1861  AGCAAGTGAGGCTTCCCACAGGCACCAGCAGCCAGCAGAATGAATGTAGCCCTTCCAACAC  1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1732  AGCAAGTGAGGCTTCCCACAGGCACCAGCAGCCAGCAGAATGAATGTAGCCCTTCCAACAC  1791

Qy   1921  CTGACAGAATGAGACCAAACGCAGCCAGTCGGAGCAAACCAAGACCATCTGAG  1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1792  CTGACAGAATGAGACCAAACGCAGCCAGTCGGAGCAAACCAAAGACCATCTGAG  1851

Qy   1981  GAATGAGAAGTCTCCGGAGGCGGCGGCCAGGGACTCTGCCGAGGAACCCCAGGGGC  2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1852  GAATGAGAAGTCTGCCGGAGGCGGCGGCCAGGGACTCTGCCGAGGAACCCCAGGGGC  1911

Qy   2041  CGAGGAGGGGGCGGGGAAGGTTCAGCGGAGTTTGCCAGAACAACCACCAGCCCCGCTCCCGCC  2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1912  CGAGGAGGGGGCGGGGAAGGTTCAGCGGAGTTTGCCAGAACAACCACCAGCCCCGCTCCCGCC  1971

Qy   2101  CCCCAGGGCTTCTGCACAGGCCATCCACTTCAGCCATCCACTCGGATCTCTCCTGAA  2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1972  CCCCAGGGCTTCTGCACAGGCCATCCACTTCAGCCATCCACTCGGATCTCTCCTCAC  2031

Qy   2161  CTCCCACGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC  2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2032  CTCCCACGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC  2091

Qy   2221  AAAATGCACACCCTTTTCTGTGCAAATCGTCTCTGTACATGTGTACATATTAGAAA  2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2092  AAAATGCACACCCTTTTCTGTGCAAATCGTCTCTGTACATGTGTACATATTAGAAA  2151

Qy   2281  GGGAAGATGTTAAGATATGTGGCCTGTGTGGGTTACACAGGGTGCCTGCAGCGCGGTAATATAT  2340
```

```
Db   2152  GGGAAGATGTTAAGATATGTGGCCTGTGTGGGTTACACAGGGTGCCTGCAGCGCGGTAATATAT  2211

Qy   2341  TTTAGAAATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTT  2400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2212  TTTAGAAATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTT  2271

Qy   2401  TTTTTCTTTTAAAGAGAAGAGCAGGCTTTCTAGAACTTTAAGACTTTAAGACTTTGGGAG  2460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2272  TTTTTCTTTTAAAGAGAAGAGCAGGCTTTCTAGAACTTTAAGACTTTAAGACTTTGGGAG  2331

Qy   2461  GTCTCACGGTGTAGAGAGAGAGAGAGTTTGAGGCCACCGCACAAAATTCACCAGAGGGAAAT  2520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2332  GTCTCACGGTGTAGAGAGAGAGAGAGTTTGAGGCCACCGCACAAAATTCACCAGAGGGAAAT  2391

Qy   2521  CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACGAAGAGGGATA  2580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2392  CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACGAAGAGGGATA  2451

Qy   2581  CCGTCTCCTTGAAGAGGAAACTCTGCTCACTCCTCATGCCTGTCTAGCTCATCACCCATT  2640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2452  CCGTCTCCTTGAAGAGGAAACTCTGCTCACTCCTCATGCCTGTCTAGCTCATCACCCATT  2511

Qy   2641  TCTCTTTGCTTCACTCTCTCCCCGCCGTTTTTGCATACTGGTTTTTGCATACTTCTCTGTCT  2700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2512  TCTCTTTTGCTTCACTCTCTCCCCGCCGTTTTTGCATACTGGTTTTTGCATACTTCTCTGTCT  2571

Qy   2701  CCTCTGTTTATCTCCCCCTCCAATCCCGTATCTCTTCTTCTCCATCTCCATCTCCATTCTTTGAA  2760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2572  CCTCTGTTTATCTCCCCCTCCAATCCCGTATCTCGTATCTCTTCTTCTCCATCTCCATTCTTTGAA  2631

Qy   2761  TTTCCTCATCCTCCAATCCCGTATCTCACGCACCCCCCCCCCCCCCCCCCCAGGCAAAGCA  2820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2632  GTGTCTGAGTATCACATCACACAAAGGAACAAAAGGAACACAAGCACAAACACAGCCTCAA  2691

Qy   2821  GTGCTCTGAGTATCACATCACACAAAGGAACAAAAGGAACAAAAGCACAAACACCAGCCTCAA  2880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2692  GTGTCTGAGTATCACATCACACAAAGGAACAAAAGGAACACAAGCACAAACACAGCCTCAA  2751

Qy   2881  CTTACACTTGGTTACTCCAAAGAACAAGAGTCAATGGTACTTGTCGTGCTTTTGGAAG  2940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2752  CTTACACTTGGTTACTCCAAAGAACAAGAGTCAATGGTACTTGTCGTGCTTTTGGAAG  2811

Qy   2941  AGGAAAACAGGAACCCACCAACCAATCAACCAAACAAAGAAAAAATTCCACAATG  3000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2812  AGGAAAACAGGAACCCACCAACCAATCAACCAAACAAAGAAAAAATTCCACAATG  2871

Qy   3001  AAAGAATGTATTTGTCTTTTTGCATTTGTATTATTCAGCATCAATATTCAGCAAAATG  3060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2872  AAAGAATGTATTTTGTCTTTTTGCATTTGTATTATTCAGCATCAATATTCAGCAAAATG  2931

Qy   3061  ATTCCTTTCTTTAAAAAAAAAAGTGGAGAAAGTAGAAATTTACCAAGGTGTTGGCC  3120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2932  ATTCCTTTCTTTAAAAAAAAAAGTGGAGAAAGTAGAAATTTACCAAGGTGTTGGCC  2991

Qy   3121  CAGGGCGTAAATTCACAGATTTTTTAACGAAGAAAAACACACAGAAGAAGCTACCTCAG  3180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2992  CAGGGCGTAAATTCACAGATTTTTTAACGAAGAAAAACACACAGAAGAAGCTACCTCAG  3051

Qy   3181  GTGTTTTTTACCTCAGCACCTTGCTCGTGTTCCCTTAGAGATTTTGTAAAGCTGAATAG  3240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3052  GTGTTTTTTACCTCAGCACCTTGCTCGTGTTCCCTTAGAGATTTTGTAAAGCTGATAG  3111

Qy   3241  TTGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAATAAGATATCAACT  3300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3112  TTGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAATAAGATATCAACT  3171

Qy   3301  GCCAGCCTGAGAAGGTGACAGTCCAAGTGTCAACAGCTGTTCTCGAATTGTCTTCCGCT  3360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3172  GCCAGCCTGAGAAGGTGACAGTCCAAGTGTCAACAGCTGTTCTCGAATTGTCTTCCGCT  3231

Qy   3361  AGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT  3412
```

```
Db    3232 AGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3283

RESULT 4
US-09-814-353-20241
; Sequence 20241, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
;   APPLICANT: Lee, John
;   APPLICANT: Thompson, Pamela
;   APPLICANT: Lillie, James
;   TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;   TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;   TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
;   FILE REFERENCE: MRI-006B
;   CURRENT APPLICATION NUMBER: US/09/814,353
;   CURRENT FILING DATE: 2001-03-21
;   PRIOR APPLICATION NUMBER: US 60/191,031
;   PRIOR FILING DATE: 2000-03-21
;   PRIOR APPLICATION NUMBER: US 60/207,124
;   PRIOR FILING DATE: 2000-05-25
;   PRIOR APPLICATION NUMBER: US 60/211,940
;   PRIOR FILING DATE: 2000-06-15
;   PRIOR APPLICATION NUMBER: US 60/216,820
;   PRIOR FILING DATE: 2000-07-07
;   PRIOR APPLICATION NUMBER: US 60/220,661
;   PRIOR FILING DATE: 2000-07-25
;   PRIOR APPLICATION NUMBER: US 60/257,672
;   PRIOR FILING DATE: 2000-12-21
;   NUMBER OF SEQ ID NOS: 22037
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20241
;   LENGTH: 3905
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1, 3897, 3898, 3899, 3900, 3901, 3902, 3903, 3904, 3905
;   OTHER INFORMATION: n = A,T,C or G
US-09-814-353-20241

Query Match      91.3%;  Score 3116.4;  DB 10;  Length 3905;
Best Local Similarity    96.1%;  Pred. No. 0;
Matches 3278;  Conservative    0; Mismatches    2; Indels 131; Gaps    3;

Qy      3 CAGCGGAGGAGGCGAGGAGGCCGGTACCGGCCGCGGGGAGCCGGGGCTCTCGGGGAA   62
          ||||||||||||||||||||||||||||||||||||||||||||  |||||||||||
Db    369 CAACGGAGGAGGCGAGGAGGCCGGTACCGGCCGCGGGGAGCCGGGGCTCTCGGGGAA  428

Qy     63 GAGACGGATGATGAACAAGCTTTACATCGGGAACTCGAGCCCCGCCGTCACCGCCGACGA  122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    429 GAGACGGATGATGAACAAGCTTTACATCGGGAACTCGAGCCCCGCCGTCACCGCCGACGA  488

Qy    123 CCTCCGGCAGCTCTTTGGGGACAGGAAGCTGCCCTGGCCCCGGCGGACAGGTCCTGCTGAAGTC  182
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    489 CCTCCGGCAGCTCTTTGGGGACAGGAAGCTGCCCTGGCCCCGGCGGACAGGTCCTGCTGAAGTC  548

Qy    183 CGGCTACGCCTTCGTCGACTACCCGACCAGGAACTGGCCGACCAGGAACATCATGGGCCATCGCC  242
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    549 CGGCTACGCCTTCGTCGACTACCCGACCAGGAACTGGCCGACCAGGAACATCATGGGCCATCGCC  608

Qy    243 CTCGGGTAAAGTGGAATTGGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAA  302
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    609 CTCGGGTAAAGTGGAATTGGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAA  668

Qy    303 GCTAAGGAGCAGGAAAATTCAGATTCAGAAACATCCCTCCTCACCTGCACCTGGGAGGTGTT  362
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    669 GCTAAGGAGCAGGAAAATTCAGATTCAGAAACATCCCTCCTCACCTGCACCTGGGAGGTGTT  728

Qy    363 GGATGCAGCTTTTGGCTCAATATGGGACAGTGGAGAATGTGGAGAATGTGGAAACTGAACAGTCAACAGACAC  422
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    729 GGATGCAGCTTTTGGCTCAATATGGGACAGTGGAGAATGTGGAGAATGTGGAAACTGAACAGACAC  788
```

(Db)   3232 ... 3283

```
Qy    423 AGAAACCGCCGTGTCAACGTCACATATGCACAAGAGAGAGAGCAAAATAGCCATGGA  482
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    789 AGAAACCGCCGTGTCAACGTCACATATGCACAAGAGAGAGAGCAAAAATAGCCATGGA  848

Qy    483 GAAGCTAAGGCGGCATCAGTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGATGA  542
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    849 GAAGCTAAGGCGGCATCAGTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGATGA  908

Qy    543 AGAGGTGAGCTCCCCTTCGCCCCCTTCGCCCCCTCAGCGGACCCAGCCACTCTTCCCGGGA  602
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    909 AGAGGTGAGCTCCCCTTCGCCCCCTTCGCCCCCTCAGCGGACCCAGCCACTCTTCCCGGGA  968

Qy    603 GCAAGGCCACGCCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTCGGAT  662
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    969 GCAAGGCCACGCCCCTGGGGGCACTTCTCAGGCCAGACAGACAGATTGATTTCCCGCTCGGAT  1028

Qy    663 CCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAACGGAGGGCTTGACCATAAAGAA  722
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1029 CCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAACGGAGGGCTTGACCATAAAGAA  1088

Qy    723 CATCACTAAGCAGAGACCCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGGAGCTGC  782
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1089 CATCACTAAGCAGAGACCCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGGAGCTGC  1148

Qy    783 AGAGAAGCTGTCACCATCCATCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATGAT  842
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1149 AGAGAAGCTGTCACCATCCATCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATGAT  1208

Qy    843 TCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAA  902
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1209 TCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAA  1268

Qy    903 AATCTTGGCACAACAATGGCCTTGGTTGGAGAACTGATTGGGAAAAGAAGGCAGAAATTTGAA  962
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1269 AATCTTGGCACAACAATGGCCTTGGTTGGTGGAACTGATTGGGAAAAGAAGGCAGAAATTTGAA  1328

Qy    963 GAAAATTGAACATGAAACAGGGACCAAGATAACAAATCTCATCTTTGCAGGATTTGAGCAT  1022
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1329 GAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCAT  1388

Qy   1023 ATACAACCCGAAAGAACCATCACTCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGA  1082
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1389 ATACAACCCGAAAGAACCATCACTCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGA  1448

Qy   1083 GATAGAGATTATGAAGAAGTGCGTGAGGCTTTGAAAATGATATGCTGCTGTTAACCA  1142
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   1449 GATAGAGATTATGAAGAAGTGCGTGAGGCCTTGAAAATGATATGCTGCTGTTAAC--  1506

Qy   1143 ACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGACT  1202
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1507 ------------------------------------------------------------  1506

Qy   1203 GTCCGTGCTATCTCCACCAGCAGCAGGGCCCGCCCCCGCGCTCCCCGCTGCCCCTACCACCC  1262
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1507 ------------------------------------------------------------  1506

Qy   1263 CTTCACTACCCACTCCGGATACTTCTCCAGCAGGAGCAGGATGTGTAACCCGTACCCCGT  1322
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1507 ------ACCCACTCCGGATACTTCTCCAGCAGGAGCAGGATTGTGTAACCCGTACCCCGTT  1559

Qy   1323 CCCGCATCATCACTCTATCCGAGCAGGAGATTGTGAATCTTCATCCAACAGGAGC  1382
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1560 CCCGCATCATCACTCTTATCCGAGCAGGAGATTGTGAATCTTCATCCAACAGGAGC  1619

Qy   1383 TGTGGGGCGCCATCATCGGGAAGAGAGGGGGCACACATCAAACAGCTGGCGGAGATTCGCCGG  1442
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1620 TGTGGGGCGCCATCATCGGGAAGAGAGGGGGCACATCAAACAGCTGGCGGAGATTCGCCGG  1679

Qy   1443 AGCTCTATCAAGATTGCCCCTGCCCGCGGAGAAATGTGGAGAATGTGGTCATCAT  1502
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1680 AGCCTCTATCAAGATTGCCCCTGCCCGCGGAGAAGCTGCAGGGAGGCGTGT  1739

Qy   1503 CACCGGGCACCACCGGAGCCAGTTCAAGCCCCAGGCCCAGGACGACGGATCTTTGGAAACTGAAAGA  1562
```

```
Db 2820 GTCTCCTTGAAGAGGAAACTCTGTCACTCCTCCTCATGCGTCCTGTCTCAGTCATACACCCATTTC 2879
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 2643 TCTTTGCTTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCT 2702

Db 2880 TCTTTGCTTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCT 2939
         ||||||||||||||||||||||||||||||||||||||||||
Qy 2703 CTCTGTTTATCTCTCCCCTCCCCTCCCCTCCCTTCCTCTCCATCTCCATTCTTTTTGAATT 2762

Db 2940 CTCTGTTTATCTCTCCCCTCCCCTCCCCTCCCCTCCCTTCCTCTCCATTCTTTTTGAATT 2999
         |||||||||||||||||||||||||||||||||||||
Qy 2763 TCCTCATCCTCCATCTCAATCCCGTATCTCACGCACCCCCCCCCAGGCAAAGCAGT 2822

Db 3000 TCCTCATCCTCCATCTCAATCCCGTATCTCACGCA-CCCCCCCCAGGCAAAGCAGT 3058
         ||||||||||||||||||||||||||||||||||||
Qy 2823 GCTCTGAGTATCACATCACACACACACACACGAAAAGCGAAACACAACCAGCCTCAACT 2882

Db 3059 GCTCTGAGTATCACATCACACACACACACACGAAAAGCGAAACACAACCAGCCTCAACT 3118
         ||||||||||||||||||||||||||||||||
Qy 2883 TACACTTGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAGAG 2942

Db 3119 TACACTTGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAGAG 3178
         |||||||||||||||||||||||||||||||||
Qy 2943 GAAAACAGGAACCCACCACAACCAATCACCAACAAAGAAAAAATTCCACAATGAA 3002

Db 3179 GAAAACAGGAACCCACCACAACCAATCACCAACAAAGAAAAAATTCCACAATGAA 3238
         |||||||||||||||||||||||||||||
Qy 3003 AGAATGTATTTTGTCTTTTTGCATTTTGGTGTATAAGCCATCAATATTCAGCAAAATGAT 3062

Db 3239 AGAATGTATTTTGTCTTTTTGCATTTTGGTGTATAAGCCATCAATATTCAGCAAAATGAT 3298
         ||||||||||||||||||||||||||||||
Qy 3063 TCCTTTCTTT-AAAAAAAAAAAAATTCACAGATTTTTTCCCTTGCTCTTGTTTTCCCTTAGAGATTTTGTAAAGCTGATAGT 3121

Db 3299 TCCTTTCTTTAAAAAAAAAAAAATTCACAGATTTTTTCCCTTGCTCTTGTTTTCCCTTAGAGATTTTGTAAAGCTGATAGT 3358
         |||||||||||||||||||||||||||||||
Qy 3122 AGGGCGGTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAGAGAGTCACTTCAGG 3181

Db 3359 AGGGCGGTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAGAGAGTCACTTCAGG 3418
         |||||||||||||||||||||||||||||||
Qy 3182 TGTTTTTACCTCAGCACCTTGCTCTTGGTCTTGTTTTTAATAAAAATGATATTAAGATATCAACTG 3241

Db 3419 TGTTTTTACCTCAGCACCTTGCTCTTGGTCTTGTTTTTAATAAAAATGAGTTGGAAAAAATAAGATATCAACTG 3478
         |||||||||||||||||||||||||||||
Qy 3242 TGGAGCATTTTTTATTTTTATTTTTAATAAAAATGAGTTGGAAAAAATAAGATATCAACTG 3301

Db 3479 TGGAGCATTTTTTATTTTTATTTTTAATAAAAATGAGTTGGAAAAAATAAGATATCAACTG 3538
         ||||||||||||||||||||||||||||||
Qy 3302 CCAGGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTCGAATTGTCTTCCGCTA 3361

Db 3539 CCAGGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCGAATTGTCTTCCCGCTA 3598
         ||||||||||||||||||||||||||||||
Qy 3362 GCCAAGAACCNATATGGCCTTCTTTTTGGACAAAACCTTGAAAATGTTTATTT 3412

Db 3599 GCCAAGAACCTATATGGCCTTCTTTTTGGACAAAACCTTGAAAATGTTTATTT 3649
```

```
Db 1740 CACCCGGCACCGAAGCCAGTTCAAGGCGACCGGATCTTTGGGAAAACTGAAAGA 1799
         |||||||||||||||||||||||||||||||||
Qy 1563 GGAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGAAGCGCATATCAGAGTGCCCTC 1622

Db 1800 GGAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGAAGCGCATATCAGAGTGCCCTC 1859
         |||||||||||||||||||||||||||||||||
Qy 1623 TTCCACAGCTGGCCGGTGATTGGCAAGACCGTGAACGAACTGCAGAACTT 1682

Db 1860 TTCCACAGCTGGCCGGTGATTGGCAAGACCGTGAACGAACTGCAGAACTT 1919
         |||||||||||||||||||||||||||||||||
Qy 1683 AACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTGAT 1742

Db 1920 AACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTGAT 1979
         ||||||||||||||||||||||||||||||||
Qy 1743 CGTCAGAATTATCGGGCACTTCTTTGCTCAGCTGCACAGGCCGCAAGATCAGGGGAAAT 1802

Db 1980 CGTCAGAATTATCGGGCACTTCTTTGCTCAGCTGCACAGGCCGCAAGATCAGGGGAAAT 2039
         |||||||||||||||||||||||||||||||
Qy 1803 TGTACAACAGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTGCGCCTCACAGCGCAG 1862

Db 2040 TGTACAACAGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTGCGCCTCACAGCGCAG 2099
         ||||||||||||||||||||||||||||||||
Qy 1863 CAAGTGAGGCTCCCACAGCCACCACGATGAATGTAGCCCTTCCAACACCT 1922

Db 2100 CAAGTGAGGCTCCCACAGCCACCACGATGAATGTAGCCCTTCCAACACCT 2159
         |||||||||||||||||||||||||||||||||
Qy 1923 GACAGAATGAGACCAAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAGA 1982

Db 2160 GACAGAATGAGACCAAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAGA 2219
         ||||||||||||||||||||||||||||||||
Qy 1983 ATGAGAAGTCTGCGGAGGCGGCCGCCAGGACTCTGCCGAGGACAACCCCGCCTCCCCCGG 2042

Db 2220 ATGAGAAGTCTGCGGAGGCGGCCGCCAGGACTCTGCCGAGGACAACCCCGCCTCCCCCGG 2279
         ||||||||||||||||||||||||||||||||
Qy 2043 AGGAGGGGCGGGGAAGGTCAGCCAGGTTGCCCAGCAGCCAGGTTGCCCAGCAAACCACCACCACTCCGAACCACCAGGGGCCG 2102

Db 2280 AGGAGGGGCGGGGAAGGTCAGCCAGGTTGCCCAGCAAACCACCACCACTCCGAACCACCAGGGGCCG 2339
         |||||||||||||||||||||||||||||||
Qy 2103 CCAGGGCTTCTGCAGCTTCGCAGCTTCAGCGCTTCAGCCATCCATCCACTTCACCGGAACTCTCCTGAACT 2162

Db 2340 CCAGGGCTTCTGCAGCTTCGCAGCTTCAGCCATCCATCCACTCCACTTCACCGGATCTCTCCTGAACT 2399
         ||||||||||||||||||||||||||||||||
Qy 2163 CCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAA 2222

Db 2400 CCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAA 2459
         |||||||||||||||||||||||||||||||
Qy 2223 AATGCACACCCTTTTCTGTGGCAAATCGTCTGTACATGTGTGTACATATTAGAAAGG 2282

Db 2460 AATGCACACCCTTTTCTGTGGCAAATCGTCTGTACATGTGTGTACATATTAGAAAGG 2519
         ||||||||||||||||||||||||||||||||
Qy 2283 GAAGATGTTAAGATATGTGGCCTGTCGGGTTACACAGGGTGCCTGCGGCGGTAATATATTT 2342

Db 2520 GAAGATGTTAAGATATGTGGCCTGTGTGGGTTACACAGGGTGCCTGCGGCGGTAATATATTT 2579
         ||||||||||||||||||||||||||||||
Qy 2343 TAGAAATAATATATCAAATAACTCAACTCCAATTTTTAATCAATTATTAATTTTTT 2402

Db 2580 TAGAAATAATATCAAATAACTCAACTAACTCCAATTTTTAATCAATTATTAATTTTTT 2639
         ||||||||||||||||||||||||||||||||
Qy 2403 TTTCTTTTTAAAGAGAAACAGCCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGT 2462

Db 2640 TTTCTTTTTAAAGAGAAACAGCCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGT 2699
         ||||||||||||||||||||||||||||||||
Qy 2463 CTCACCGTGTAGAGAGACAGCTTTGAGCCACCCGACACAAAATTCACCCGAGGGGAAATCT 2522

Db 2700 CTCACCGTGTAGAGAGACAGCTTTGAGCCACCCGACCACAAAATTCACCCGAGGGGAAATCT 2759
         ||||||||||||||||||||||||||||||
Qy 2523 CGTCGGAAGGACACTCACGGCCAGTTCTCGGATCACCTGTATGTCAACAGAAGGGATACC 2582

Db 2760 CGTCGGAAGGACACTCACGGCCAGTTCTGGATCACCTGTATGTCAACAGAAGGGATACC 2819
         ||||||||||||||||||||||||||||||
Qy 2583 GTCTCCTTGAAGAGGAAACTCTGTCACTCCTCCTCATGCGTCCTGTCTCAGTCATACACCCATTTC 2642
```

```
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
;   Assessment, Prevention, and Therapy of Ovarian Cancer
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 146
; LENGTH: 3667
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-097-340-146

Query Match          91.3%;   Score 3114.8;   DB 15;   Length 3667;
Best Local Similarity 96.1%;  Pred. No. 0;
Matches 3277;  Conservative 0;  Mismatches 0;  Indels 131;  Gaps 3;

Qy    3 CAGCGGAGGAGGCGAGGAGCGCGGTACCGGCCCGGGGAGCCCGGGGCTCTCGGGAA 62
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db  369 CAACGGAGGAGGCGAGGAGCGCGGTACCGGCCCGGGGAGCCCGGGGCTCTCGGGAA 428

Qy   63 GAGACCGATGATGAACAAGCTTTACATCGGGAAACTTACATCGGGAAACTGAACCGA 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  429 GAGACCGATGATGAACAAGCTTTACATCGGGAAACTTACATCGGGAAACTGAACCGA 488

Qy  123 CCTCCCGGACAGCTCTTTGGGACAGGAAGCTGCCCTGGCGGGACAGGTCCTGCTGAAGTC 182
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  489 CCTCCCGGACAGCTCTTTGGGACAGGAAGCTGCCCTGGCGGGACAGGTCCTGCTGAAGTC 548

Qy  183 CGGCTACGCCTTCGTGACTACCCCAGACCCCAGAGAGACCCT 242
          |||||||||||||||||||||||||   |||||||||||||||
Db  549 CGGCTACGCCTTCGTGACTACCCCAGACCCCAGAGAGACCCT 608

Qy  243 CTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGGAAGTTGATTACTCAGTCTCTAAAA 302
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  609 CTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGGAAGTTGATTACTCAGTCTCTAAAA 668

Qy  303 GCTAAGGAGCAGGAAAATTCAGATTCGAAATTCGAAACATCCCTCCTCACCTGCAGTGGTGTT 362
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  669 GCTAAGGAGCAGGAAAATTCAGATTCGAAATTCGAAACATCCCTCCTCACCTGCAGTGGTGTT 728

Qy  363 GGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGAACAGTCAACACAGACAC 422
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  729 GGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGAACAGTCAACACAGACAC 788

Qy  423 AGAAAACCGCCGGTTGTCACGTCACATATGCAACAGAAGAAGCAAAAATAGCCATGGA 482
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  789 AGAAAACCGCCGGTTGTCACGTCACATATGCAACAGAAGAAGCAAAAATAGCCATGGA 848

Qy  483 GAAGCTAAGCGGGCATCCCCTTCGCCCCCTCAGCGCGGAGCCCAGCCGTGGGGA 542
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  849 GAAGCTAAGCGGGCATCCCCTTCGCCCCCTCAGCGCGGAGCCCAGCCGTGGGGA 908

Qy  543 AGAGGTGAGCTCCCCTCGCCCCCTCGCCCCCTCAGCGCGGAGCCCAGCCGTGGGGA 602
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  909 AGAGGTGAGCTCCCCTCGCCCCCTCGCCCCCTCAGCGCGGAGCCCAGCCGTGGGGA 968

Qy  603 GCAAGGCCACGCCCTGGGGGCACTTCTCAGGCGCAGCCAGACAGATGATTCCGCTGCGGAT 662
          ||||||||||||||||||||||||||||||||   |||||||||||||||||||||||||
Db  969 GCAAGGCCACGCCCTGGGGGCACTTCTCAGGCCCAGACAGATTGATTCCGCTGCGGAT 1028

Qy  663 CCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGGAAAAGGAGGGCTTGACCATAAAGAA 722
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1029 CCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGGAAAAGGAGGGCTTGACCATAAAGAA 1088

Qy  723 CATCACTAAGCAGACCCAGTCCCGGGTAGTATCCATAGAAAGAGAACTCTGGAGCTGC 782
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1089 CATCACTAAGCAGACCCAGTCCCGGGTAGTATCCATAGAAAGAGAACTCTGGAGCTGC 1148

Qy  783 AGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCCATGAT 842
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1149 AGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCCATGAT 1208

Qy  843 TCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAA 902
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1209 TCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAA 1268

Qy  903 AATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGAAGGCAGAAATTTGAA 962
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1269 AATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGAAGGCAGAAATTTGAA 1328

Qy  963 GAAAATTGAACATGAAAACAGGGACCAAGATAACAATCTCATCTTTGCAGATTTGAGCAT 1022
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1329 GAAAATTGAACATGAAAACAGGGACCAAGATAACAATCTCATCTTTGCAGATTTGAGCAT 1388

Qy 1023 ATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGCCTGGCTGGCTGTTAACCA 1082
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1389 ATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGCCTGGCTGGCTGTTAACCA 1448

Qy 1083 GATAGAGATTATGATGAAGAAGCTGCGTGAGGGCTTTGAAAATGATATGATATGCTGGCTGTTAAC-- 1506
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1449 GATAGAGATTATGATGAAGAAGCTGCGTGAGGGCTTTGAAAATGATATGATATGCTGGCTGTTAAC 1506

Qy 1143 ACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGGCACTTGAATCTCTTTCAACAGGGACT 1202
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1507 ------------------------------------------------------------- 1506

Qy 1203 GTCCGTGCTATCTCCACCAGCAGCAGGGGCCCCGCGGAGCTCCCCGCGCCCTACCACCC 1262
          |||||||||||||||||||||||||||||||||
Db 1507 ------------------------------------------------------------- 1506

Qy 1263 CTTCACTACCCACTCCGGATACTTCCCAGCCTGCTACCCCATCACCAGTTTGGCCCGTT 1322
          |||||||||||||||||||   ||||||||||||||||||||||||||||||
Db 1507 ------ACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGCCCGTT 1559

Qy 1323 CCCGCATCATCACTCTTATCCAGACAGCAGGGGGGCACAATCAAACAGTCTGGGGAAGAAGAGGGGG 1382
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1560 CCCGCATCATCACTCTTATCCAGACAGCAGGGGGGCACAATCAAACAGTCTGGGGAAGAAGAGGGGC 1619

Qy 1383 TGTGGGCGCCATCATCGGGAAGAACATCAGTCTGGGCGGAAAGGATGGTCATCAT 1442
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1620 TGTGGGCGCCATCATCGGGAAGAACATCAGTCTGGGCGGAAAGGATGGTCATCAT 1679

Qy 1443 AGCCTCTATCAAGATTGCCCCGCGGAAGCCCAGATTCTCAGCGAACGTCAGCGAAAGACAC 1502
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1680 AGCCTCTATCAAGATTGCCCCGCGGAAGCCCAGATTCTCAGCGAACGTCAGCGAAAGACAC 1739

Qy 1503 CACCGGGCCACCCAGGCCCAGTTCAAGGCCCAGTTCTTGGGAAACTGAAAGA 1562
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1740 CACCGGGCCACCCAGGCCCAGTTCAAGGCCCAGTTCTTGGGAAACTGAAAGA 1799

Qy 1563 GGAAAACTTCTTTAACCCCAAAGAAGTTCCTACATCCCGATGA 1622
          |||||||||||||||||||||||||||||||||||||||||||
Db 1800 GGAAAACTTCTTTAACCCCAAAGAAGTTCCTACATCCCGATGA 1859

Qy 1623 TTCCACAGCTGCCGCGGTGATTGGCAAGACCGTGCAAGAACCTGCAGAACTT 1682
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1860 TTCCACAGCTGCCGCGGTGATTGGCAAGACCGTGCAAGAACCTGCAGAACTT 1919
```

```
Qy   1683  AACCAGTGCAGAAGTCATCGTCGCTCCTGACCAAACGCCAGATGAAATGAGGAAGTGAT  1742
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1920  AACCAGTGCAGAAGTCATCGTCGCTCCTGTGACCAAACGCCAGATGAAATGAGGAAGTGAT  1979

Qy   1743  CGTCAGAATTATCGGGACACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAAT  1802
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1980  CGTCAGAATTATCGGGACACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAAT  2039

Qy   1803  TGTACAACAGGTGAAGCAGCAGGAGCAGAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAG  1862
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2040  TGTACAACAGGTGAAGCAGCAGGAGCAGAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAG  2099

Qy   1863  CAAGTGAGGCTCCCACAGGGACCAGCAGCAAACAACGATGAATGTAGCCCTTCCAACACCT  1922
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2100  CAAGTGAGGCTCCCACAGGGACCAGCAGCAAACAACGATGAATGTAGCCCTTCCAACACCT  2159

Qy   1923  GACAGAATGAGACCAAACGCCAGCCAGACCAGTCGGGAGCAAACCAAAGACCATCTGAGGA  1982
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2160  GACAGAATGAGACCAAACGCCAGCCAGACCAGTCGGGAGCAAACCAAAGACCATCTGAGGA  2219

Qy   1983  ATGAGAAGTCTGCGGAGGCGGGCAGGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGCCG  2042
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2220  ATGAGAAGTCTGCGGAGGCGGGCAGGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGCCG  2279

Qy   2043  AGGAGGGGCGGGGGAAGCTCAGCCAGGGTTTGCCAGAACCACCAGGATCCTCCTGAACT  2102
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2280  AGGAGGGGCGGGGGAAGCTCAGCCAGGGTTTGCCAGAACCACCAGGATCCTCCTGAACT  2339

Qy   2103  CCAGGGGCTTCGCAGGCTTCAGCCATCCACCTCACCTTGCCTGTTCACCACTCACCTCCTGAACT  2162
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2340  CCAGGGGCTTCGCAGGCTTCAGCCATCCACCTCACCTTGCCTGTTCACCACTCACCTCCTGAACT  2399

Qy   2163  CCCACGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAA  2222
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2400  CCCACGCACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAA  2459

Qy   2223  AATGCACACCCTTTTCTGTGGCAAATGTCTCTGTACATGTGTGTACATATTAGAAAGG  2282
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2460  AATGCACACCCTTTTCTGTGGCAAATGTCTCTGTACATGTGTGTACATATTAGAAAGG  2519

Qy   2283  GAAGATGTTAAGATATGTGGCCTGTGTGGTTACACAGGGTTACCAGCCGTAATATATTT  2342
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2520  GAAGATGTTAAGATATGTGGCCTGTGTGGTTACACAGGGTTACCAGCCGTAATATATTT  2579

Qy   2343  TAGAAATAATATATCAAATAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTATTTTT  2402
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2580  TAGAAATAATATATCAAATAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTATTTTT  2639

Qy   2403  TTTCTTTTTAAAGAGAAAGCAGAGCCTTTCTTTTGAGACTTTAAAGAATAAAGTCTTTGGGAGT  2462
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2640  TTTCTTTTTAAAGAGAAAGCAGAGCCTTTCTTTTGAGACTTTAAAGAATAAAGTCTTTGGGAGT  2699

Qy   2463  CTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGCACAAATTCACCCAGAGGGAAATCT  2522
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2700  CTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGCACAAATTCACCCAGAGGGAAATCT  2759

Qy   2523  CGTCGGAAGGACACTCACGGCAGTTCTGGATCACGTGTATGTCAACAGAAGGGATACC  2582
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2760  CGTCGGAAGGACACTCACGGCAGTTCTGGATCACGTGTATGTCAACAGAAGGGATACC  2819

Qy   2583  GTCTCCTTGAAGAGAAACTCTGTCACTCCTCATGCCTGTCTAGCTCATCACCCCATTTC  2642
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2820  GTCTCCTTGAAGAGAAACTCTGTCACTCCTCATGCCTGTCTAGCTCATCACCCCATTTC  2879

Qy   2643  TCTTTTGCTTCACCAGTTTTTAAACTGGTTTTTTGCATACGTGTTTTTTGCATACTGCTATATAATTCTCTGTCTCT  2702
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2880  TCTTTTGCTTCACCAGTTTTTAAACTGGTTTTTTGCATACGTGTTTTTTGCATACTGCTATATAATTCTCTGTCTCT  2939

Qy   2703  CTCTGTTTATTCTCTCCCCTTCTTCCCCTTCTTCTCTCCCTTCTCTCCATCTCCATTCTTTTTGAATT  2762
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2940     CTCTGTTTATTCTCTCCCCTTCTTCCCCTTCTTCTCTCCCTTCTCTCCATCTCCATTCTTTTTGAATT  2999

Qy   2763  TCCTCATCCCTCCATCTCAATCCCGTATCTACGGCACCCCCCCCCCCCGTCACCGACCCCCCCCAGGCAAAGCAGT  2822
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   3000  TCCTCATCCTCCATCTCCAATCTCCGTATCTCACGGCA-CCCCCCCCCCCAGGCAAAGCAGT  3058

Qy   2823  GCTCTGAGTATCACATCACACAAAAGGAACACAAAGCGAAACACACAAACCAGCCTCAACT  2882
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3059  GCTCTGAGTATCACATCACACAAAAGGAACACAAAGCGAAACACACAAACCACCAGCCTCAACT  3118

Qy   2883  TACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAG  2942
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3119  TACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAG  3178

Qy   2943  GAAAACAGGAACCCACCAAACCAAACAACCAATCAACCAAACAAAGAAAAATTCCACAATGAA  3002
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3179  GAAAACAGGAACCCACCAAACCAAACAACCAATCAACCAAACAAAGAAAAATTCCACAATGAA  3238

Qy   3003  AGAATGTATTTGTCTTTTGCATTTGTGTATAAGCCATCAATATTCAGCAAAATGAT  3062
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3239  AGAATGTATTTGTCTTTTGCATTTGTGTATAAGCCATCAATATTCAGCAAAATGAT  3298

Qy   3063  TCCTTTCTTT-AAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCCC  3121
            |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   3299  TCCTTTCTTTAAAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCCC  3358

Qy   3122  AGGGCGTTAAATTCACAGATTTTTTAACGAGAAAACACACAGAAGAAGCTACCTCAGG  3181
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3359  AGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAAGAAGCTACCTCAGG  3418

Qy   3182  TGTTTTACCTCAGCACCTTGCTCCTTGTTCTTAATAAAAAAATGAGTTGGAAAAAAAATAAGATATCAACTG  3241
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3419  TGTTTTACCTCAGCACCTTGATCCTTGTTCTTAATAAAAAAATGAGTTGGAAAAAAAATAAGATATCAACTG  3478

Qy   3242  TGGAGCAATTTTTTATTTTTTATTTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACTG  3301
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3479  TGGAGCAATTTTTTATTTTTTATTTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACTG  3538

Qy   3302  CCAGCCTGGAGAAGGTGACAGTCCAAGTGTCAACAGCTGTTCGAATTGTCTTCCCGCTA  3361
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3539  CCAGCCTGGAGAAGGTGACAGTCCAAGTGTCAACAGCTGTTCGAATTGTCTTCCGCTA  3598

Qy   3362  GCCAAGAACCNATATGGCCTTCTTTTGACAAACCTTGAAAATGTTTATTT  3412
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db   3599  GCCAAGAACCTATATGGCCTTCTTTTGACAAACCTTGAAAATGTTTATTT  3649
```

```
RESULT 6
US-10-648-593-48
; Sequence 48, Application US/10648593
; Publication No. US20040106132A1
; GENERAL INFORMATION:
;    APPLICANT: Bristol-Myers Squibb Company
;    TITLE OF INVENTION: IDENTIFICATION OF GENES FOR PREDICTING ACTIVITY OF COMPOUNDS THAT
;    TITLE OF INVENTION: INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE KINASES AND/OR
;    TITLE OF INVENTION: PROTEIN TYROSINE KINASE PATHWAYS IN BREAST CELLS
;    FILE REFERENCE: D0273 NP
;    CURRENT APPLICATION NUMBER: US/10/648,593
;    CURRENT FILING DATE: 2003-08-26
;    PRIOR APPLICATION NUMBER: 60/406,385
;    PRIOR FILING DATE: 2002-08-27
;    NUMBER OF SEQ ID NOS: 557
;    SOFTWARE: PatentIn version 3.2
; SEQ ID NO 48
;    LENGTH: 3667
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-10-648-593-48

  Query Match             91.3%;  Score 3114.8;  DB 17;  Length 3667;
  Best Local Similarity   96.1%;  Pred. No. 0;
  Matches 3277;  Conservative    0;  Mismatches    3;  Indels  131;  Gaps     3;

Qy      3  CAGCGGAGGAGGCGAGGAGCGCGGGGCTACCCGGGGGAGCCGCGGGGGAGCCGCGGGGGCTCTCGGGGAA   62
            || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    369  CAACGGAGGAGGCGGAGGAGCGCGGGGCTACCCGGGGGAGCCGCGGGGGAGCCGCGGGGGCTTCGGGGAA  428
```

This page is a dense nucleotide sequence alignment printout comparing two sequences labeled "Qy" (query) and "Db" (database). The alignment is presented in paired rows with position numbers flanking each line. The left-hand block and right-hand block together form a continuous alignment.

Representative readable position numbering (Qy / Db rows):

Qy start	Qy end	Db start	Db end
63	122	429	488
123	182	489	548
183	242	549	608
243	302	609	668
303	362	669	728
363	422	729	788
423	482	789	848
483	542	849	908
543	602	909	968
603	662	969	1028
663	722	1029	1088
723	782	1089	1148
783	842	1149	1208
843	902	1209	1268
903	962	1269	1328
963	1022	1329	1388
1023	1082	1389	1448
1083	1142	1449	1506
1143	1202	1507	1506
1203	1262	1507	1506
1263	1322	1507	1559
1323	1382	1560	1619
1383	1442	1620	1679
1443	1502	1680	1739
1503	1562	1740	1799
1563	1622	1800	1859
1623	1682	1860	1919
1683	1742	1920	1979
1743	1802	1980	2039
1803	1862	2040	2099
1863	1922	2100	2159
1923	1982	2160	2219
1983	2042	2220	2279
2043	2102	2280	2339
2103	2162	2340	2399
2163	2222	2400	2459
2223	2282		

```
Db  3539 CCAGCCTGGAGGAAGTGCAGTCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCTA 3598
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3362 GCCAAGAACCNATATGGCCTTCTTTGGACAAACCTTGAAAATGTTTATTT 3412
                ||||||||| ||||||||||||||||||||||||||||||||||||||||
Db  3599 GCCAAGAACCTATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3649

RESULT 7
US-09-764-864-332
; Sequence 332, Application US/09764864
; Patent No. US2002013752753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 332
;   LENGTH: 2290
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-864-332

Query Match       58.3%;     Score 1989.8;     DB 9;     Length 2290;
Best Local Similarity 99.4%;  Pred. No. 0;
Matches 2028;  Conservative 0;  Mismatches 8;  Indels 8;  Gaps 3;

Qy  1376 CCCAGGCTGTGGGCGCCATCATCGGGAGGAAGGGGGCACACATCAAACAGCTGGCGAGAT 1435
                |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db    25 CCCACGCTGTCCGGCGCCATCATCGGGAGGAAGGGGGCACACATCAAACAGCTCGGCGAGAT 84

Qy  1436 TCGCCGGAGCCTCTATCAAGATGCCCCTGCGGAAGGCCCAGAGCGTCAGCGGAAAGGATGG 1495
                ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db    85 TCGCCGGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCGTCAGCGGAAAGGATGG 144

Qy  1496 TCATCATCACCGGGCCACCGGAAGCCCAGTTCAAGGGCCCAGGGACGGATCTTTGGGAAAC 1555
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   145 TCATCATCACCGGGCCACCGGAAGCCCAGTTCAAGGGCCCAGGGACGGATCTTTGGGAAAC 204

Qy  1556 TGAAAGAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGGAAGCGCATATCAGAG 1615
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   205 TGAAAGAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGGAAGCGCATATCAGAG 264

Qy  1616 TGCCCCTTCCACAGCTGCGCCGGGTGATTGGCAAGGTGGCAAGACCGGTGAACGAACTGC 1675
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   265 TGCCCCTTCCACAGCTGCGCCGGGTGATTGGCAAGGTGGCAAGACCGGTGAACGAACTGC 324

Qy  1676 AGAACTTAACCAGTGCAGAAGAATTATCGGGCACTCTTCTTCTTTGCTAGCGACCGCAGAAATGAGG 1735
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   325 AGAACTTAACCAGTGCAGAAGAATTATCGGGCACTCTTCTTCTTTGCTAGCGACCGCAGAAATGAGG 384

Qy  1736 AAGTGATCGTGACAACAGGTGAAGCAGCAGCACCAGCAGACCAGCAGAAATACCCTCAGGGAGTCGCCTTCAC 1795
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   385 AAGTGATCGTGACAACAGGTGAAGCAGCAGCACCAGCAGACCAGCAGAAATACCCTCAGGGAGTCGCCTTCAC 444

Qy  1796 GGGAAATTGTACAACAGGTGAAGCAGCAGAGCAGCAGGACCACCAGGATGAATGTAGCCCTTCC 1855
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   445 GGGAAATTGTACAACAGGTGAAGCAGCAGAGCAGCAGGACCACCAGGATGAATGTAGCCCTTCC 504

Qy  1856 AGCCGAGCAAGTGAGGCTTCCCACAGGCTCCCACAGCCAGCAAACAACGGCGGGAGCAAACGACCAT 1915
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   505 AGCCGAGCAAGTGAGGCTTCCCACAGGCTCCCACAGCCAGCAAACAACGGCGGGAGCAAACGACCAT 564

Qy  1916 AACACCTGACAGAATGAGACCAAAGCCAGCTGCGGGAGCAAACCAAAGACCAT 1975
                ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   565 AACACCTGACAGAATGAGACCAAAGCCAGCTGCGGGAGCAAACCAAAGACCAT 624

Qy  1976 CTGAGGAATGAGAAGTCTGCGGAGGCGGCCGAGGCCCTGAGAACCCCA 2035
                ||||||||||||||||||||||||||||||||||||||||||||||||
Db   625 CTGAGGAATGAGAAGTCTGCGGAGGCGGCCGAGGCCCTGAGAACCCCA 684
```

Left column:

```
Db  2460 AATGCACACCCTTTTCTGTGGCAAATCGTCTCGTACATGTGTACATATTAGAAAAGG 2519
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2283 GAAGATGTTAAGACATGTGGGCCTGTGGTTACACAGGGTGCTGCCTGCGCGGTAATATATT 2342
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2520 GAAGATGTTAAGACATGTGGGCCTGTGGTTACACAGGGTGCTGCCTGCGCGGTAATATATT 2579

Qy  2343 TAGAAATAATATCAAATAACTCAACTAACTCCAATTTTTAATCAATTATTAATTTTTT 2402
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2580 TAGAAATAATATCAAATAACTCAACTAACTCCAATTTTTAATCAATTATTAATTTTTT 2639

Qy  2403 TTTCTTTTTAAAGAGAAAGCAGGCTTTCTAGACTTTAAAGAATAAAGTCTTGGGAGGT 2462
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2640 TTTCTTTTTAAAGAGAAAGCAGGCTTTCTAGACTTTAAAGAATAAAGTCTTGGGAGGT 2699

Qy  2463 CTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGACAAAATTCACCCAGAGGGAAATCT 2522
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2700 CTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGACAAAATTCACCCAGAGGGAAATCT 2759

Qy  2523 CGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAGAGGGATACC 2582
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2760 CGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAGAGGGATACC 2819

Qy  2583 GTCTCCTTGAAGAGGAAACTTCTGTCACTCCTCATGCCGTGTCTAGCTCATCACCCATTTC 2642
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2820 GTCTCCTTGAAGAGGAAACTTCTGTCACTCCTCATGCCGTGTCTAGCTCATCACCCATTTC 2879

Qy  2643 TCTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCTCT 2702
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2880 TCTTTGCTTCACAGGTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCTCT 2939

Qy  2703 CTCTGTTTATCTCTCCCCTCCCCCTCCCCCTTCTTCTCCATCCTCCATCCCATTCTCTTTGAATT 2762
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2940 CTCTGTTTATCTCTCCCCTCCCCCTCCCCCTTCTTCTCCATCCTCCATCCCATTCTCTTTGAATT 2999

Qy  2763 TCCTCATCCCTCCAATCCGTATCTCACGCACCCCCCCCCCCCCCCCCCCAGGCAAAGCAGT 2822
               |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db  3000 TCCTCATCCCTCCAATCCGTATCTCACGCA-CCCCCCCCCCCCCAGGCAAAGCAGT 3058

Qy  2823 GCTCTGAGTATCACATCACCACAAAGGAACAAAAGCGAAACACACAAACCAGCCTCAACT 2882
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3059 GCTCTGAGTATCACATCACCACAAAGGAACAAAAGCGAAACACACAAACCAGCCTCAACT 3118

Qy  2883 TACACTTGGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAGAG 2942
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3119 TACACTTGGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAGAG 3178

Qy  2943 GAAAACAGGAACCCCACCAACCAACCAACAAAGAAAATTCCACAATGAA 3002
               ||||||||||||||||||||||||||||||||||||||||||||||||
Db  3179 GAAAACAGGAACCCCACCAACCAACCAACAAAGAAAATTCCACAATGAA 3238

Qy  3003 AGAATGTATTTGTCTTTTGCATTTGGTGTGAAAGTGTGGAGAAAGTAGAATTACCAAGGTCAAAATGAT 3062
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3239 AGAATGTATTTGTCTTTTGCATTTGGTGTGAAAGTGTGGAGAAAGTAGAATTACCAAGGTCAAAATGAT 3298

Qy  3063 TCCTTTCTTT-AAAAAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCCC 3121
               |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db  3299 TCCTTTCTTTAAAAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCCC 3358

Qy  3122 AGGGCGTTAAATTCACAGATTTTTTTTGCTTCGTTCTTGTTTCCCTGTGTTCTGTTTAAGGAGCTACCTCAGG 3181
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3359 AGGGCGTTAAATTCACAGATTTTTTTTTAACGAGAAAAACACACAGAAGAGAGCTACCTCAGG 3418

Qy  3182 TGTTTTTACCTCAGCACCTTGTCTTGCTTCTTGTTTCCCTGAGATTTTGTAAAGCTGATAGT 3241
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3419 TGTTTTTACCTCAGCACCTTGATCTTGTTTCCCTAGAGATTTTGTAAAGCTGATAGT 3478

Qy  3242 TGGAGCATTTTTTTTATTTTTTTAATAAAATGAGTTGGAAAAATAAGATATCAACTG 3301
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3479 TGGAGCATTTTTTTTTATTTTTTTAATAAAATGAGTTGGAAAAATAAGATATCAACTG 3538

Qy  3302 CCAGCCTGGAGAAGTGCAGTCAACGTGTCAACAGCTGTTCTTCCGCTA 3361
               ||||||||||||||||||||||||||||||||||||||||||||||||
Db  3362 CTGAGGAATGAGAAGTCTGCGGAGGCGGCCGAGGCCCTGAGAACCCCA
```